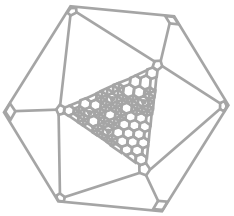
University of Kent

SCHOOL OF BIOSCIENCES

Characterisation of the Fucose and Rhamnose Utilising Bacterial Microcompartment system in *Clostridium phytofermentans*



A thesis submitted to the University of Kent for the degree of PhD in the Faculty of Sciences

Shouaa A. Alrobaish

2019

Declaration

I hereby declare that this thesis written presents my own work which has been done after registration for the degree of *Doctor of Philosophy* (*PhD*) of the Department of Bioscience of The University of Kent.

This Thesis, or any equivalent or similar form, has not been submitted or presented to any other university in support for an application for any degree, diploma or any other qualification, other than the university, for which I am now a candidate.

Abstract

Bacterial microcompartments (BMCs) are cytoplasmic polyhedral organelles that are composed of a semi-permeable proteinaceous shell made from a number of different proteins, which encapsulate a specific metabolic process. Recent studies have revealed that the basic functions of these compartments are not only to increase the metabolic efficiency of the enclosed enzymes but also to protect the cell from toxic intermediates. BMCs are found in about 20 % of all bacteria and spread widely across the kingdom. One of the bacteria that contains operons encoding for such protein compartments is *Clostridium phytofermentans*, a bacterium that was originally isolated from forest soil and can directly convert plant biomass into biofuels.

The bacterial genome of *C. phytofermentans* has three BMC-encoding loci, one of which is a 13-gene operon that houses the genetic information for a predicted fucose/rhamnose utilisation system. Within this operon, six genes encode for shell proteins, which form the BMCs outer structure, and the other seven genes encode for metabolic enzymes that are responsible for converting fucose/rhamnose into ethanol, propanol, and propionate. Based on this genome analyses, research has been undertaken to clone the various BMC-associated genes of the *C. phytofermentans* fucose/rhamnose system in order to generate a recombinant BMC in *E. coli*. This was attempted firstly by the individual cloning and recombinant expression in *E. coli* of the genes that encode the shell proteins. This approach also allowed the characterisation of the shell proteins and determine their ability to form functional BMCs. Secondly, engineer empty BMCs from the six shell proteins and investigate the minimum number of genes required for BMC formation.

One of the shell proteins characterised in this study, Cphy_1186 (PduT-like), was crystallised and had its structured determined by X-ray crystallography. The protein was found to have a trimeric arrangement of subunits, each of which contains a double-BMC-domain consistent with it belonging to the BMC-T class. The trimer forms a flat approximately hexagonally shaped disc with a central pore that is suitable for binding a 4Fe-4S cluster. The pore is lined with six cysteine residues, more than enough for Fe-S centre formation

The structure of the shell protein Cphy_1182 (PduA-like) was also solved. The protein contains a single-BMC-domain within each subunit and forms a hexamers, thereby belonging to the BMC-H class. Overproduction of Cphy_1182 by itself results in the formation of filaments or nanotubes in the bacterial cell.

The first attempt at engineering empty metabolosomes from the six shell proteins that are involved in the formation of the *C. phytofermentans* metabolosome is presented. Although ultimately this was unsuccessful the associated research has given greater insight into the structure/function relationship between the individual components.

Acknowledgments

Firstly and foremost, I would like to express my gratitude to Prof. Martin Warren and Prof. Mark Smales, my research supervisors, for giving me the opportunity to be part of their research groups as well as for their patient guidance, enthusiastic encouragement and useful critiques of this research work.

Special thanks to also go to Dr. Stefanie Frank for her kindness, support and encouragement, especially for her help in developing my lab skills and having confidence in my ability. You were always next to me at the time I was most in need and I am eternally grateful.

My grateful thanks are extended to all members of the Warren lab for helping me during my PhD study. I would particularly like to thank Dr. Matthew Lee and Maria Stanley for their support inside the lab and for assistance in proof reading this thesis, Marie Anderson for taking my protein crystals for X-ray crystallography and Dr. Mingzhi Liang for his support inside the lab and his help in preparing samples for TEM analysis. A past member of the lab whom I must also thank is Dr. David Palmer who showed me how to use the Belle Technology glove-box for anaerobic purification and crystallization. He also taught me how to prepare solutions, collect and store samples anaerobically.

Special thanks go to Prof. David Brown for explaining how to solve proteins structures and for proof reading the crystallography chapter. I also give

thanks to Jack Bradley (Queen Mary, UoL) for helping me to determine the structure of the Cphy_1186 shell protein.

I would like to express my very great appreciation to Dr. Charlotte Harrison for her kindness and support, and for sharing her office with me whilst I was writing up, as well as for help in improving my English through conversation and many interesting discussions

Special thanks should be given to Ian Brown for his support as he provided me with my initial TEM training and improved my image analysis understanding.

I would also like to extend my thanks to Dr. Stephen Rigby (University of Manchester) for EPR analysis and explaining the results.

Last but not the least, I would like to thank my family. Particularly my husband who encouraged me not to change from studies in biology during my time as a postgraduate, supported me in many aspects, and always told me how proud he is of my achievements. During my time in Kent my parents have been a major source of emotional support and have provided a happy family atmosphere through their daily call. My brothers and sister have supported me spiritually throughout the writing of this thesis and my life in general.

Table of contents

Declaration…… **II**

Abstract…..…… **III**

Acknowledgments…… **IV**

Table of contents…….. **VI**

Abbreviation…….. **XII**

List of Figures…… **XV**

List of Tables……. **XX**

Chapter 1 – Introduction

1.1 Introduction to bacterial microcompartments…….…… 22

1.2 Carboxysomes……..…… 26

1.2.1 Carboxysome shell proteins……..…… 31

1.2.1.1 CcmK……..……. 32

1.3 Metabolosomes……..……. 33

1.3.1 Propanediol utilisation……..….. 34

1.3.1.1 Reactivation and recycling of diol dehydratase..... 39

1.3.1.2 Pdu shell proteins……..…… 40

1.4 Ethanolamine utilisation……..…….. 44

1.5 Fucose and rhamnose utilisation by *C. phytofermentans*….. 48

1.5.1 *C. phytofermentans* fucose and rhamnose fermentation pathways……..…….. 51

1.6 Aims and objectives……..…….. 53

Chapter 2 – Materials and Methods

2.1 Materials…… 55

2.1.1 Chemicals…… 55

2.1.2 Bacterial strains…… 55

2.1.3 Plasmids…… 56

2.1.4 Primers…… 57

2.1.5 Media and solutions for culturing…… 58

2.1.6 Media and solutions for DNA procedures…… 61

2.1.7 Media and solutions for protein procedures…… 62

2.1.7.1 Solutions for protein purification…… 62

2.1.7.2 Solutions for microcompartment purification…… 63

2.1.7.3 Solutions for SDS-PAGE…… 64

2.1.7.4 Solutions for EPR experiments…… 66

2.1.7.5 Solutions for Transmission electron microscopy experiments…… 67

2.2 Microbiological methods…… 68

2.2.1 Sterilisation…… 68

2.2.2 Bacterial storage…… 69

2.2.3 Solid cultures…… 69

2.2.4 Liquid cultures…… 69

2.2.5 Preparation of competent cells…… 70

2.2.6 Transformation of competent cells…… 70

2.2.7 Production of recombinant protein…… 71

2.2.8 Lysis of cells using sonication…… 71

2.3 Molecular biological methods…… 71

2.3.1 PCR reactions…… 71

2.3.2 DNA Electrophoresis……. 72

2.3.2.1 Agarose gel……. 72

2.3.2.2 Visualisation under UV……. 73

2.3.3 DNA extraction……..……...…… 73

2.3.3.1 Extraction from Agarose gel…….……... 73

2.3.3.2 Extraction from solutions…….…....….... 73

2.3.3.3 Extraction of plasmid DNA……..……..……..... 74

2.3.4 Restriction enzyme digest……..……..…… 74

2.3.5 DNA ligation……. 75

2.3.6 Cloning into vectors……..……..…… 75

2.3.6.1 Cloning into pGEM®-T Easy-Indirect cloning…..…. 75

2.3.6.2 Cloning into pET3a, pET14b and pLysS vectors.... 76

2.3.6.3 Multi-cloning strategy for the *C. phytofermentans* Cphy genes…….…….....….. 77

2.3.6.4 Mutagenesis by overlap extension of *Cphy_1186*.. 79

2.4 Biochemical methods……...….. 80

2.4.1 Standard purification procedures…….……. 80

2.4.1.1 Immobilised metal affinity chromatography…..... 80

2.4.2 Buffer exchange…….…..…….. 84

2.4.3 Calculation of protein concentration…….…… 84

2.4.4 Fast protein liquid chromatography…….……..…. 84

2.4.5 Polyacrylamide gel electrophoresis…….……...……. 85

2.4.5.1 SDS-PAGE……..……...…….……. 85

2.4.6 Anaerobic techniques…….…… 85

2.4.7 Microcompartment purification…….……..…. 86

2.4.8 Preparation of samples for transmission electron microscopy…… 87

2.4.9 Sectioning and staining of samples…….…….... 88

2.4.10 Electron paramagnetic resonance spectroscopy……. 88

2.4.11 Protein crystallisation……..……. 89

Chapter 3 – Characterisation of the *C. phytofermentans* fucose and rhamnose bacterial microcompartment shell proteins

3.1 Introduction…….…….. 93

3.2 Results…….……... 94

3.2.1 Sequence analysis…….……... 94

3.2.2 Amplification and cloning of genes encoding the six shell proteins……. 97

3.2.3 Shell protein overproduction, purification and characterisation……... 100

3.2.3.1 Characterisation and solubilisation of Cphy_1176….. 101

3.2.3.2 Characterisation of Cphy_1180…….…. 105

3.2.3.3 Characterisation of Cphy_1181……. 109

3.2.3.4 Characterisation of Cphy_1182……. 116

3.2.3.5 Characterisation of Cphy_1184……. 120

3.2.3.6 Characterisation of Cphy_1186……. 124

3.2.4 Engineering empty bacterial microcompartment Shell……... 128

3.3 Conclusion…….…… 131

Chapter 4 – Study of formation of overproduced recombinant *Cphy* proteins

4.1 Introduction…….…….. 133

4.2 Results…….……... 134

4.2.1 Expression of individual shell proteins…….…….. 134
4.2.1.1 Characterisation of Cphy_1176……. 134
4.2.1.2 Characterisation of Cphy_1180……. 137
4.2.1.3 Characterisation of Cphy_1181……. 139
4.2.1.4 Characterisation of Cphy_1182……. 140
4.2.1.5 Characterisation of Cphy_1184……. 143
4.2.1.6 Characterisation of Cphy_1186……. 144
4.2.2 Expression of engineered empty BMCs……..……. 147
4.2.3 Purification of semi-BMCs structures…… 149
4.2.4 Effect of different combinations of shell proteins on organelle formation……. 151
4.2.4.1 Expression of Cphy_1176-1180……. 152
4.2.4.2 Expression of Cphy_1176-1180-1181…….. 154
4.2.4.3 Expression of Cphy_1176-1180-1181-1182…… 156
4.2.4.4 Expression of Cphy_1176-1180-1181-1182-1184….. 159
4.2.4.5 Expression of Cphy1184-1176……..….. 161
4.2.4.6 Expression of Cphy1184-1186…….…….. 162
4.3 Conclusion..…… 164

Chapter 5 – Protein crystallisation and structure analysis

5.1 Introduction…….…….. 167
5.2 Results……... 168
5.2.1 Structure analysis of Cphy_1181……... 168
5.2.1.1 Crystallisation of Cphy_1181……. 169
5.2.2 Structure analysis of Cphy_1182……... 173
5.2.2.1 Crystallization of Cphy_1182……. 174
5.2.2.2 Data collection and molecular replacement……..… 175
5.2.3 Structure analysis of Cphy_1186……... 184
5.2.3.1 Anaerobic purification of Cphy_1186…… 186

5.2.3.2 Crystallization of Cphy_1186……. 188
5.2.3.3 Data collection…….. 190
5.2.3.4 Molecular replacement and subunit analysis……... 193
5.2.3.5 Trimeric structure…… 196
5.2.3.6 Molecular tiling of Cphy_1186……... 200
5.2.3.7 Iron sulfur [4Fe-4S] binding site……..…..… 203
5.2.3.8 Mutagenesis……. 204
5.2.3.9 Electron Paramagnetic Resonance (EPR)……... 207
5.3 Conclusion…….…….. 211

Chapter 6 – Discussion

6.1 Discussion……..……..…… 214

Images of controls

Control section of *E. coli* containing pET3a (induced overnight with 200 µM IPTG)…… 221
Control section of *E. coli* containing pLysS (induced overnight with 200 µM IPTG)…… 221
Control section of *E. coli* containing pET3a (induced for 2 h with 400 µM IPTG)……..….. 222
Control section of *E. coli* containing pET3a (induced for 2 h with 200 µM IPTG)……. 222
Control of Figure 4.14……. 223

References

References……...……..… 224

Abbreviations

1,2-PD	1,2-Propanediol
Å	Ångstrom
Ado	Adenosyl
Ado-B_{12}	Adenosylcobalamin
AIM	Autoinduction media
bp	Base pairs
BMC	Bacterial microcompartment
cbbL	Large subunit of IA RuBisCO that related to α-carboxysomes
cbbS	Small subunit of IA RuBisCO that related to α-carboxysomes
Ccm	Carbon dioxide concentrating mechanism
CcmK1	β-carboxysomes shell protein 1
CcmK2	β-carboxysomes shell protein 2
CcmK3	β-carboxysomes shell protein 3
CcmK4	β-carboxysomes shell protein 4
CoA	Coenzyme A
CO_2	Carbon dioxide
Cphy	*Clostridium phytofermentans* gene
Cphy_1176	Metabolosome shell protein
Cphy_1180	Metabolosome shell protein
Cphy_1181	Metabolosome shell protein
Cphy_1182	Metabolosome shell protein
Cphy_1184	Metabolosome shell protein
Cphy_1186	Metabolosome shell protein
CsoS1	α-Carboxysome shell
CsoS2	α-Carboxysome scaffold protein
Da	Dalton
dH_2O	Distilled water
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
e.g.	*Exempli gratia* (for example)

EPR	Electron paramagnetic resonance
Eut	Ethanolamine utilisation
EutA	Coenzyme B_{12} assimilation & recycling
EutB	Ethanolamine ammonia-lyase large subunit (Ethanolamine degradation)
EutC	Ethanolamine ammonia-lyase small subunit (Ethanolamine degradation)
EutD	Ethanolamine degradation protein
EutE	Aldehyde dehydrogenase (Ethanolamine degradation)
EutG	Ethanolamine degradation protein
EutH	Ethanolamine facilitator
EutK	Ethanolamine shell protein
EutL	Ethanolamine shell protein
EutM	Ethanolamine shell protein
EutN	Ethanolamine shell protein
EutR	Transcription activator
EutS	Ethanolamine shell protein
Fe	Iron
Fe-S	Iron-sulfur
FPLC	Fast protein liquid chromatography
HCl	Hydrochloric acid
HCO3 -	Bicarbonate
His	Histidine
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β-D-1- thiogalactopyranoside
k	Kilo
kb	Kilo base
kDa	Kilo Dalton
LB	Lysogeny broth
NAD^+	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide (reduced form)
Ni	Nickel
nm	Nanometre
OH	Hydroxyl
OD	Optical density
PCR	Polymerase chain reaction
PDB	Protein data bank

PEG	Polyethylene glycol
Pdu	1,2-propanediol utilisation
rbcL	Large subunit of IB RuBisCO that related to β-carboxysomes
rbcS	Small subunit of IB RuBisCO that related to β-carboxysomes
RuBisCO	Ribulose 1,5-bisphosphate carboxylase
RuBP	Ribulose-1,5-bisphosphate
SDS-PAGE	Sodium dodecyl sulfate – polyacrylamide gel electrophoresis
TEM	Transmission electron micrograph
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
UV-Vis	Ultraviolet visible

List of Figures

Figure 1.1 The bacterial microcompartment and associated shell proteins. 24

Figure 1.2 Outer representation of the crystallised compete closed shell of unknown function from *Haliangium ochraceum*. 25

Figure 1.3 A model for carboxysome BMC where CO_2 sequestered in order to enhance the activity of RuBisCO. 27

Figure 1.4 TEM analysis of cells expressing carboxysomes. 28

Figure 1.5 Crystal packing of Ccmk shell proteins in molecular layers. 33

Figure 1.6 TEM analysis of the Pdu microcompartment. 35

Figure 1.7 Schematic of the 24 gene pdu operon of *Salmonella* responsible for BMC assembly and 1,2-PD degradation. 37

Figure 1.8 Model for the Pdu metabolosome where 1,2-PD degradation pathway occurs inside the lumen to optimise growth and confining volatile metabolic intermediates (propionaldehyde). 38

Figure 1.9 TEM image of purified *C. freundii* PduA shell protein recombinantly expressed in *E. coli*. 41

Figure 1.10 The three suggested models for PduA-derived nanotubes. 42

Figure 1.11 A diagram showing the relationship between shell proteins of the microcompartments**.** 44

Figure 1.12 TEM analysis of of *S. enterica* (strain JE8392 eut+) grown in 30 mM ethanolamine. 45

Figure 1.13 Schematic of the 17 gene *eut* operon of *Salmonella* responsible for BMC assembly and ethanolamine degradation. 46

Figure 1.14 A model for the Eut metabolosome where ethanolamine degradation happens to optimise growth and confining volatile metabolic intermediates (acetaldehyde). 47

Figure 1.15 Tree of *C. phytofermentans* and other strains from 7 different clusters. 49

Figure 1.16 TEM analysis of *C. phytofermentans* ISDg. 50

Figure 1.17 *C. phytofermentans* fucose and rhamnose fermentation pathways based on sequence homology and expression data. 52

Figure 2.1 5 µL HyperLadder™ 1kb run on a 1 % (w/v) Agarose gel in TAE. 61

Figure 2.2 Pre-stained Blue Protein Standard, Broad Range is a mixture of highly purified, prestained proteins, covalently coupled with a blue chromophore that resolves into 11 sharp bands when electrophoresed**.** 66

Figure 2.3 Schematic drawing of 'Link and Lock' cloning. 78

Figure 2.4 Schematic drawing of mutagenesis technique. 79

Figure 2.5 Schematic drawing of cell lysis technique. 82

Figure 2.6 Schematic drawing of protein purification technique (IMAC). 83

Figure 2.7 Schematic drawing of crystallisation phase diagram. 90

Figure 2.8 Schematic drawing of the hanging drop method. 91

Figure 3.1 C. phytofermentans ISDg (Accession number NC_010001) fucose/ rhamnose utilisation 13-genes operon without the acetate kinase Cphy1327 and Cphy1187 the transcriptional regulator from DeoR family. 93

Figure 3.2 Gel electrophoresis of PCR products of the six shell proteins Cphy1176, Cphy1180, Cphy1181, Cphy1182, Cphy1184 and Cphy1186. 98

Figure 3.3 pET-3a vector including cloning, expression, and T7 N-terminal Tag region. 99

Figure 3.4 pET-14b vector including cloning, expression, and T7 N-terminal Tag, His Tag region. 99

Figure 3.5 The pGEM®-T Vector is derived from the pGEM®-5Zf(+) Vector. 100

Figure 3.6 Sequence analysis for *C. phytofermentans* Cphy_1176 shell protein compared with PduU. 102

Figure 3.7 15 % SDS-PAGE gel showing the IMAC purification of Cphy_1176 shell protein. 102

Figure 3.8 15 % SDS-PAGE gel showing soluble (S) and insoluble (P) fractions of overexpressed Cphy_1176 shell protein. 103

Figure 3.9 15 % SDS-PAGE gel showing IMAC & PD-10 purification process fractions of soluble (S), insoluble (P), total (T), binding buffer (BB), washing buffer I (WBI), washing buffer II (WBII) and elution (E) of overexpressed Cphy_1176 shell protein. 104

Figure 3.10 Sequence analysis for *C. phytofermentans* Cphy_1180 shell protein compared with PduA. 105

Figure 3.11 IMAC purification of Cphy_1180. 107

Figure 3.12 15 % SDS-PAGE gel of fractions from Cphy_1180 shell protein purification. 108

Figure 3.13 Aerobic Purification and gel filtration of Hexa-His tagged Cphy_1180 protein (pduA-like). 109

Figure 3.14 IMAC purification of Cphy_1181. 111

Figure 3.15 15 % SDS-PAGE gel photo of Cphy_1181 purification. 112

Figure 3.16 Purified proteins treated by Bio-rad protein reagent. 113

Figure 3.17 Photo of the Cphy_1181 (PduK-like) transformation from precipitated to soluble protein. 114

Figure 3.18 Further purification of Cphy_1181 via gel filtration chromatography. 115

Figure 3.19 IMAC purification of Cphy_1182. 117

Figure 3.20 15 % SDS-PAGE gel of Cphy_1182 shell protein purification. 118

Figure 3.21 Cphy_1182 gel filtration chromatography curve. 119

Figure 3.22 Sequence analysis for *C. phytofermentans* Cphy_1184 shell protein compared with EutN. 120

Figure 3.23 IMAC and PD-10 purification of Cphy_1184 protein. 121

Figure 3.24 15 % SDS-PAGE gel of Cphy_1184 protein IMAC as well as PD-10 purification processes. 122

Figure 3.25 UV-visible spectrum of aerobically purified Cphy_1184 shell protein in 20 mM Tris-HCl (pH 8.0), with 100 mM NaCl. 123

Figure 3.26 Further purification of Cphy_1184 by gel filtration chromatography. 123

Figure 3.27 IMAC and PD-10 purification of Cphy_1186 protein. 125

Figure 3.28 15 % SDS-PAGE gel of Cphy_1186 protein IMAC as well as PD-10 purification processes. 126

Figure 3.29 UV-visible spectrum of aerobically purified Cphy_1186 shell protein in 20 mM Tris-HCl (pH 8.0), with 100 mM NaCl. 127

Figure 3.30 Further purification of Cphy_1186 by gel filtration chromatography. 127

Figure 3.31 Y-PER™ purification process of the recombinant BMCs. 129

Figure 3.32 15 % SDS-PAGE gel showing purification process. 130

Figure 4.1 Sequence alignment of *C. phytofermentans* Cphy_1176 and *S. enterica* PduU. 135

Figure 4.2 Thin section of *E. coli* overproducing Cphy_1176 (PduU-like) protein. 136

Figure 4.3 Multiple sequence alignment of Cphy_1180, Cphy_1182, *C. freundii* PduA, *S. enterica* PduA. 137

Figure 4.4 Thin section of E. coli overproducing Cphy_1180 (PduA-like) protein. 138

Figure 4.5 Multiple sequence alignment of Cphy_1181 with Cphy_1182 and PduK proteins from different organisms. 139

Figure 4.6 Transmission electron microscopy of native Cphy_1181 (PduK-like) protein produced in *E. coli*. 140

Figure 4.7 Transmission electron microscopy of *E. coli* cells expressing native Cphy_1182 (PduA-like). 142

Figure 4.8 Multiple sequence alignment of Cphy_1184 with PduN and EutN proteins from *Salmonella*. 143

Figure 4.9 Transmission electron micrograph of *E. coli* cells expressing native Cphy_1184 (PduN-like). 144

Figure 4.10 Sequence alignment of Cphy_1186 with PduT shell protein from *Salmonella*. 145

Figure 4.11 Transmission electron microscopy of *E. coli* cells expressing native Cphy_1186 (PduT-like). 146

Figure 4.12 Transmission electron micrographs of *E. coli* cells expressing the engineered 148

constructs.

Figure 4.13 Thin sectioned and stained samples following purification. 150

Figure 4.14 Negative stain TEM of purified empty semi-BMCs structures the red arrow indicates one of the structures. 151

Figure 4.15 TEM analysis of *E. coli* cells expressing Cphy_1176-1180 with condition of 200 µM IPTG overnight. 153

Figure 4.16 Transmission electron micrographs of *E. coli* cells expressing Cphy_1176-1180-1181 with condition of 200 µM IPTG overnight. 155

Figure 4.17 Transmission electron micrographs of *E. coli* expressing Cphy_1176-1180-1181-1182 with condition of 200 µM IPTG overnight. 157

Figure 4.18 Transmission electron micrographs of *E. coli* cells expressing Cphy_1176-1180-1181-1182 with condition of 200 µM IPTG overnight. 158

Figure 4.19 Effects of combination of Cphy_1184-1180-1181-1182-1184 for the *Cphy* microcompartments formation. 160

Figure 4.20 TEM analysis of *E. coli* cells expressing Cphy_1184-1176 for the *Cphy* microcompartment. 161

Figure 4.21 TEM analysis of *E. coli* cells expressing Cphy_1184-1186 for the *Cphy* microcompartments. 163

Figure 5.1 Cphy_1181 shell protein 100 amino acid sequence. 169

Figure 5.2 Optimisation of Cphy_1181 crystallisation conditions. 171

Figure 5.3 Optimisation of Cphy_1181 crystallisation conditions. 172

Figure 5.4 Hanging drop of Cphy_1181 (PduK-like) protein crystals. 173

Figure 5.5 Hanging drop of Cphy_1182 protein showing different sizes of crystals. 175

Figure 5.6 A Ramachandran plot generated from *C. phytofermentans* Cphy_1182, a dimer protein that contains both β-sheet and α-helix. 176

Figure 5.7 Cartoon representation of the tertiary structure of Cphy_1182 dimer which comprises four Zn. 177

Figure 5.8 Surface representation of the generated hexamer of Cphy_1182. 178

Figure 5.9 Crystal packing of Cphy_1182 dimers. 180

Figure 5.10 A model of the zinc atom interact a Histidine and two Glutamic residues. 181

Figure 5.11 Two Cphy_1182 subunits showing the edge residues Arg-157, Pro-158, Asp-128 and Asn-107. 182

Figure 5.12 A surface representation of Cphy_1182 hexamer showing the sides and the edge residues of the hexamer. 183

Figure 5.13 The subunit pore of Cphy_1182 showing the positions of residues that line the pore. 183

Figure 5.14 Cphy_1186 shell protein 182 amino acid sequence. 185

Figure 5.15 Anaerobic purification technique of Cphy_1186. 187

Figure 5.16 UV-visible absorbance spectrum of anaerobic isolated Hexa-His tagged Cphy_1186 protein. 187

Figure 5.17 Optimised conditions of Cphy_1186 anaerobic crystallisation method. 189

Figure 5.18 Anaerobic hanging drop of Cphy_1186 (PduT-like) protein crystal. 190

Figure 5.19 Cphy_1186 crystal shooting. 192

Figure 5.20 Electron density fitting of Cphy_1186 structure. 194

Figure 5.21 The tertiary structure of the Cphy_1186 subunit. 195

Figure 5.22 A quaternary and higher order structures involve in Cphy_1186. 197

Figure 5.23 A quaternary structure of Cphy_1186 contains three atoms of Zn and one atom of Na. This side is the convex side of the trimer. 198

Figure 5.24 The subunit pore measurements of Cphy_1186 showing the positions of Cysteine residues that line the pore. 199

Figure 5.25 The crystal structure packing of Cphy_1186. 201

Figure 5.26 The crystal structure packing of Cphy_1186. 202

Figure 5.27 Cphy_1186 and the binding site for 4Fe-4S. 204

Figure 5.28 Mutagenesis of Cphy_1186 shell protein 182 amino acid sequence. 205

Figure 5.30 15 % SDS PAGE for mutated Cphy_1186 IMAC fractions. 206

Figure 5.31 UV-visible absorbance spectrum of two Cphy_1186 mutations that purified anaerobically. 207

Figure 5.32 X-band CW EPR spectra of Cphy_1186 anaerobically purified. 209

Figure 5.33 X-band CW EPR spectra of Cphy_1186 expressed in whole cells with and without reduction showing [4Fe-4S] signal. 210

List of Tables

Table 2.1 List of bacterial strains used in this project. 55

Table 2.2 List of vectors used in this project. 56

Table 2.3 List of designed primers used in this project. 57

Table 2.4 Antibiotics used for bacterial resistance. 60

Table 2.5 Gel composition for SDS-PAGE and 2D-PAGE. 65

Table 2.6 Basic PCR reaction protocol and values. 72

Table 2.7 Physical PCR reaction protocol. 72

Table 2.8 Typical restriction enzymes digestion protocol. 74

Table 2.9 Typical DNA ligation protocol. 75

Table 2.10 Promega Quik pGEM®-T Easy vector ligation protocol. 76

Table 3.1 Some related information of *C. phytofermentans* ISDg genome. 94

Table 3.2 DNA and amino acids sequences of each shell protein in *C. phytofermentans*. 95

Table 3.3 genetic organisation and some functional equivalent of *C. phytofermentans* fucose/ rhamnose bacterial microcompartment. 96

Table 3.4 Ten different conditions were used to solubilise the shell protein Cphy_1176. 103

Table 3.5 Peaks and calibration values of Cphy_1180 shell protein. 109

Table 3.6 Peaks and calibration values of Cphy_1181 shell protein with and without Hexa-His tag. 115

Table 3.7 Peaks and calibration values of Cphy_1182 shell protein. 119

Table 3.8 Peaks and calibration values of Cphy_1184 shell protein. 124

Table 3.9 Peak and calibration values of Cphy_1186 shell protein. 128

Table 5.1 Protein concentrations used in crystallography screening to optimise the best condition. 170

Table 5.2 Cphy_1186 sequence analysis for cysteine detection. 185

Table 5.3 Crystallographic data statistics for native Cphy_1186. 191

Table 5.4 EPR samples of Cphy_1186 Fe-S cluster analysis. 208

Table 6.1 Summary of the shell protein characteristics. 219

CHAPTER 1

INTRODUCTION

1.1 Introduction to bacterial microcompartments

Organelles are generally a defining characteristic of eukaryotic cells, where they are used to compartmentalize metabolic processes and to separate genetic information from the cytoplasm. However, prokaryotic cells can also form internal cellular organelles called bacterial microcompartments (BMCs), which sequester specific metabolic activities. Bioinformatic studies indicate that around 20 % of bacterial species possess the genetic information that encodes for these proteinaceous organelles (Axen *et al*., 2014). BMCs are spread widely across bacterial phyla, where 40 % of bacteria producing BMCs are observed to have multiple BMC gene clusters (Sturms *et al.*, 2015). With the rapid growth in the number of sequenced bacterial genomes new BMCs will very likely be discovered (Cheng *et al.*, 2008).

The BMCs were initially discovered in the 1950s and were termed carboxysomes. Carboxysomes play an important role in carbon fixation and were observed in the cytoplasm of cyanobacteria as well as chemoautotrophic bacteria by electron microscopy (Shively *et al.*, 1973; Chowdhury., *et al.*, 2014).

In 1999 and after decades of carboxysome discovery, a new type of BMC, that compartmentalises a different metabolic process, was discovered in *Salmonella enterica* (Bobik *et al.*, 1999). Further bioinformatic and experimental work has revealed a diverse range of BMCs, which are associated with catabolic metabolic processes across the eubacterial kingdom. For instance, 1,2-propanediol utilisation (Pdu BMC) and ethanolamine utilisation (Eut BMC). These BMCs, termed metabolosomes,

are composed of structural proteins which are paralogs of the carboxysome shell proteins (Kofoid *et al.*, 1999). Since then, many more BMCs systems have been found through sequencing studies and in 2014, Axen *et al* developed a comprehensive taxonomy programme called LoClass in order to enable the identification and prediction of functional (sub) type of BMCs. Based on the analysis of sequences in the Non-redundant Protein Database (NR) of the NCBI and by comparing the BMC loci of seven candidate bacterial phyla, they were able to identify 23 different type of BMCs (including the previously identified carboxysomes and metabolosomes).

The basic function of these compartments is not only to increase the metabolic efficiency of the enclosed enzymes, but also to protect the cell from toxic intermediates (Chowdhury *et al.*, 2014). BMCs vary in size but are generally 40 – 200 nm in diameter. Evidence indicates that BMCs consist entirely of a semi-permeable shell that encapsulates sequential metabolic enzymes (Cheng *et al.*, 2008). They are formed of facets, each of which is composed of a number of shell proteins **(Figure 1.1)**. Shell proteins have a unique conserved protein domain (protein folding unit) known as a BMC domain, which consists of approximately 100 amino acid residues. The 3-dimentional structure of these individual shell proteins is composed of a β-α-β fold pattern although some shell proteins contain two of these motifs that are connected by a β-hairpin (Kerfeld *et al.*, 2005; Yeates *et al.*, 2010; Sargent *et al.*, 2013).

Studies indicate that there are three basic types of shell proteins containing a BMC-domain. The first is a protein with a single copy of the conserved BMC domain that forms a circular hexameric disk and is known as BMC-H. The

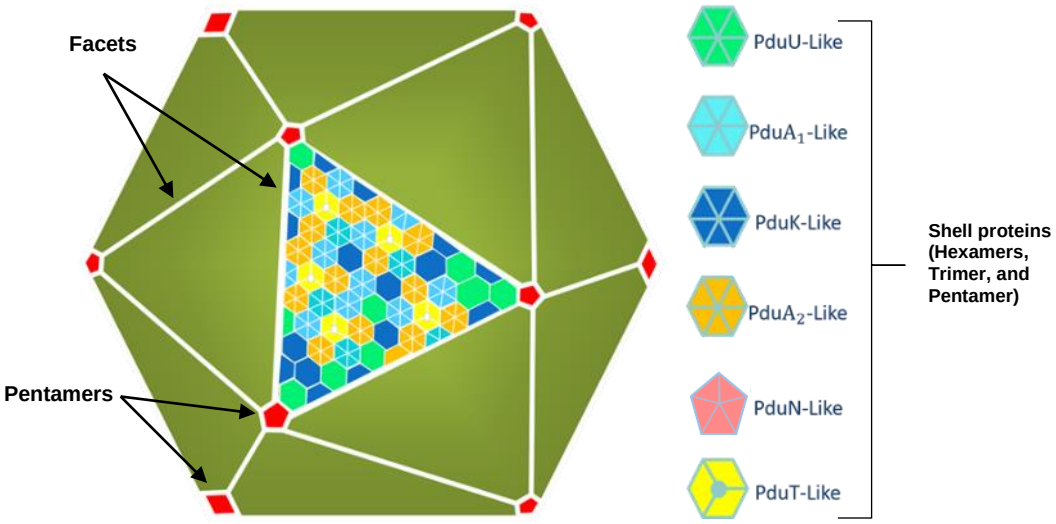


Figure 1.1: The bacterial microcompartment and associated shell proteins. This structure example of BMC was created according to *Clostridium phytofermentans* fucose and rhamnose utilisation locus.

second type of shell protein is one that contains two copies of the same conserved BMC domain (Pfam00936) and which forms a psedohexameric trimer. This is referred to as BMC-T (Cai *et al*., 2009). The final type is a protein that forms pentamers and is known as BMC-P and contains the BMC domain Pfam03319 (Zarzycki *et al*., 2015).

Recently, new investigations have revealed the basic design principles of BMC construction and shell protein organization through the crystallisation of a complete recombinant BMC of unknown function from *Haliangium ochraceum* (Lassila *et al.*, 2014; Sutter *et al.*, 2017). The recombinant production of this BMC involved the expression of a hexamer, a pentamer as well as three trimers. The recombinant BMC was purified, crystallised, and analysed by x-ray crystallography and also cryo-electron microscopy (Sutter *et al.*, 2017). The resultant BMC was found to be 40 nm in diameter and composed of 60 hexamers (BMC-H), 20 trimers (BMC-T) and 12 pentamers

(BMC-P) **(Figure 1.2)**. Within this structure, four different interfaces are observed. This includes two hexamer-hexamer interactions, one in a planar orientation and the other with a 30° tilt. The other interactions include a hexamer-pentamer interaction as well as a hexamer-trimer interaction that's bent at 30° and 25° respectively. Furthermore, the structure revealed information concerning the orientation of the shell proteins as well as the presence of double stacking trimers. The structure revealed that the concave surface of the hexamers and trimers, including the double layered trimers, always face outwards. The orientation of the pentamer was also revealed. The pentamer has two sides, one which is broader and relatively flat that

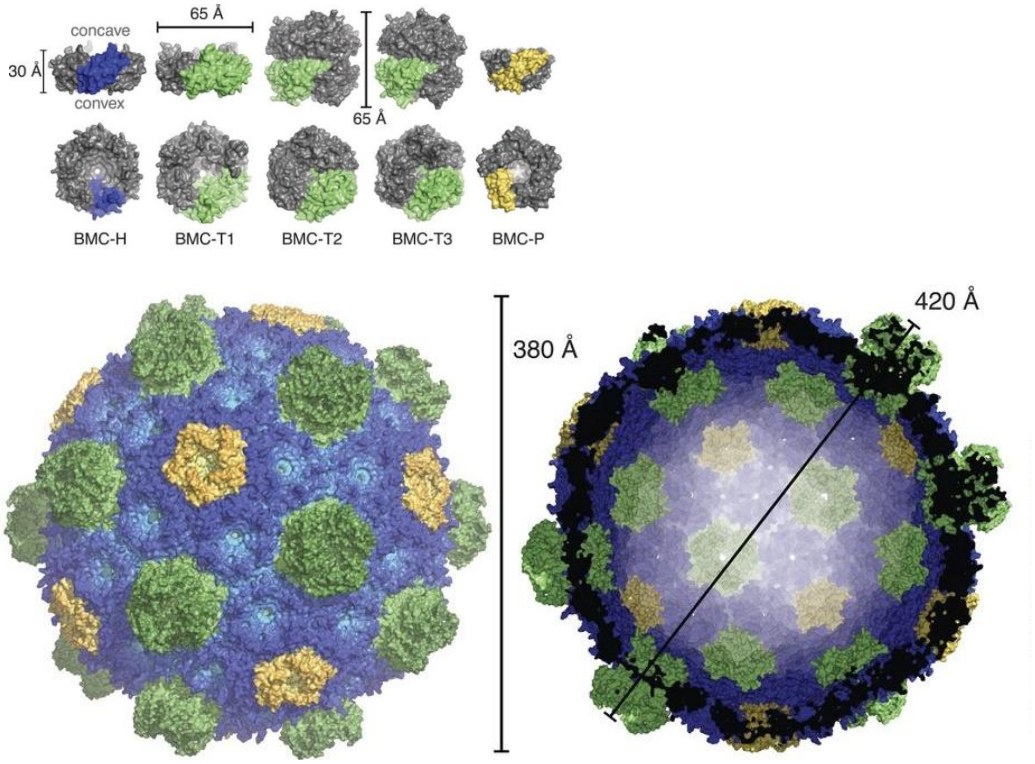


Figure 1.2: Outer representation of the crystallised compete closed shell of unknown function from *Haliangium ochraceum* (Sutter *et al*., 2017).

faces outwards, and the other which is domed and faces the lumen (Sutter *et al.*, 2017).

1.2 Carboxysomes

The first observed BMC was the carboxysome. These polyhedral structures, which are approximately 100 to 200 nm in diameter, contain the enzymes carbonic anhydrase as well as ribulose bisphosphate carboxylase (RuBisCO) and are located in the cytoplasm of cyanobacteria and chemoautotrophic bacteria (Chowdhury *et al*., 2014). It's been suggested that the carboxysome shell is roughly 2 to 3 nm thick (Kerfeld *et al.*, 2005). Recently a similar thickness was reported for the crystal structure analysis of the intact BMC shell from *Haliangium ochraceum* (Sutter *et al.*, 2017). It has been estimated that carboxysomes are responsible for approximately 25 % of global CO_2 fixation (Rae *et al.*, 2013). They were first isolated in 1973 and it was at this stage that they were shown to contain RuBisCO (Shively *et al.*, 1973). Moreover, it was observed that carboxysome production, and RuBisCO activity, increased proportionally when cyanobacteria as well as chemoautotrophic bacteria were placed in a low carbon dioxide medium. Collectively, this indicated that carboxysomes are bacterial organelles that perform carbon dioxide fixation (Beudeker *et al*. 1980; Miller *et al*.1984). The carboxysome is an essential part of the CO_2 concentrating mechanism (CCM) and is used to improve the efficiency of RuBisCO (Price *et al*., 2008). Carboxysomes are approximately 60 % RuBisCO, which is one of the most abundant proteins on earth, with each organelle containing around 300 molecules of RuBisCO (Cannon and Shively, 1983).

RuBisCO is involved in the first stage of carbon fixation in the Calvin cycle and catalyses the reaction of RuBP with CO_2. Typically, each molecule of RuBisCO fixes between 3 and 10 molecules of CO_2 every second (Ellis., 2010). RuBisCO is very sensitive to O_2 which acts as an alternative substrate to CO_2 and reduces the efficiency of the energy in terms of carbon fixation. Therefore, organisms that employ RuBisCO have to protect the enzyme from O_2, which in bacteria is achieved by the use of carboxysomes.

The process of carbon fixation starts when the bicarbonate diffuses from the cytoplasm into the lumen of the carboxysome via pores in shell proteins. The enzyme carbonic anhydrase converts bicarbonate into CO_2 which is then fixed by RuBisCO **(Figure 1.3).** It is thought that CO_2 is unable to diffuse out of the carboxysome leading to a high local concentration which enhances

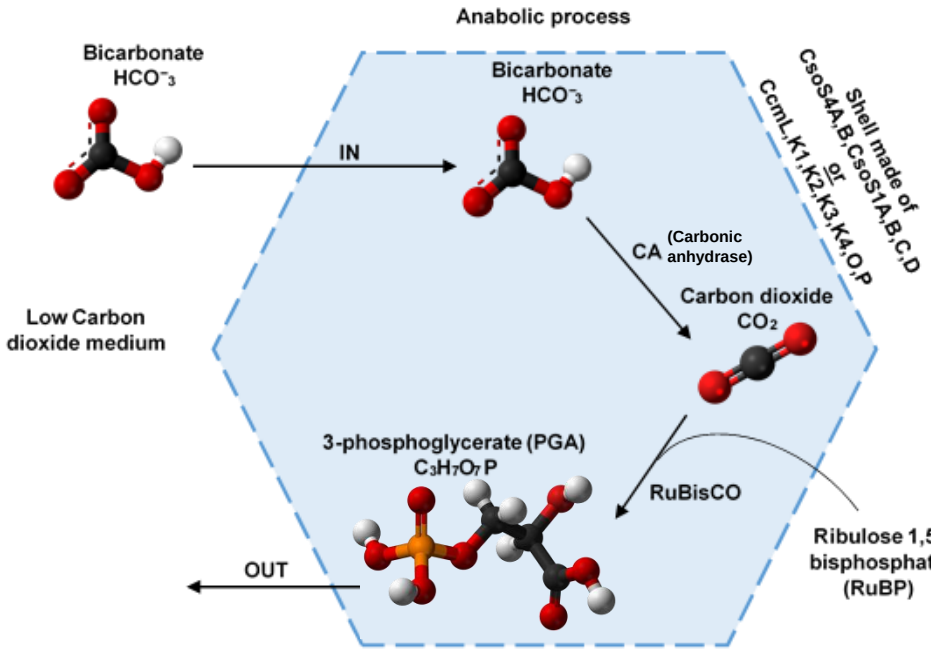


Figure 1.3: A model for carboxysome BMC where CO_2 sequestered in order to enhance the activity of RuBisCO.

CO_2 fixation and reduces competing photorespiration, which is a non-productive process that utilises O_2 and wastes up to 50 % of the fixed carbon (Chowdhury, *et al.*, 2014).

Based on sequence analysis and gene organization two classes of carboxysomes have been identified the α and β-carboxysome. α-carboxysomes are found in Prochlorococcus, some marine *Synechococcus* species and in some chemoautotrophic bacteria, while β-carboxysomes are found in all other cyanobacteria (Tripp *et al*., 2010). α-carboxysomes, which contain the form IA RuBisCO, and also referred to as the Cso type of carboxysome, and the β-carboxysome which contains form IB RuBisCO that has the Ccm (named for carbon-concentrating mechanism) locus **(Figure 1.4)** (Robert Tabita, 1999; Badger and Bek, 2008). In cyanobacteria, the genes for β-carboxysome in *Synechococcus* strain PCC7942 are located in the Ccm cluster, where the structural proteins are encoded immediately upstream from the RuBisCO large (rbcL) and small (rbcS) subunit genes. In

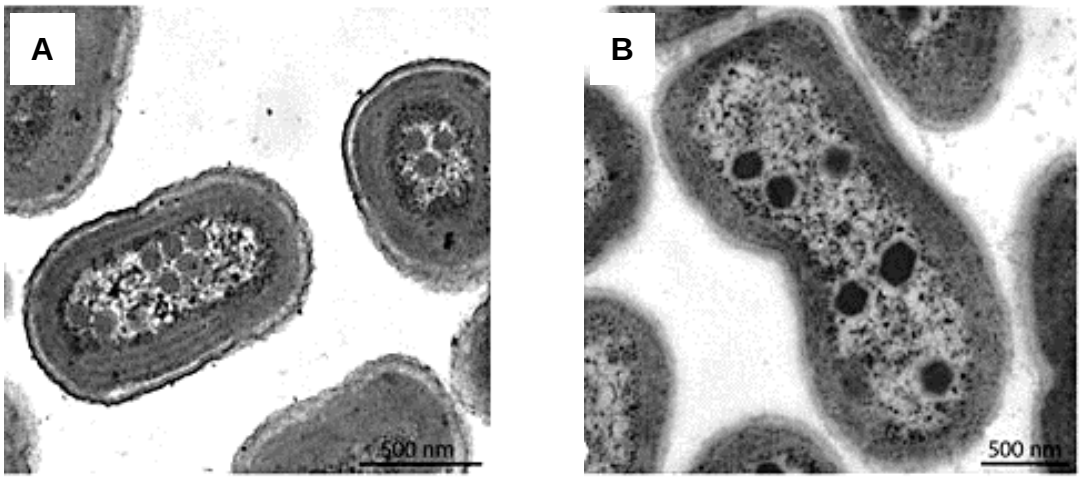


Figure 1.4: TEM analysis of cells expressing carboxysomes. (A) Thin section of *Cyanobium* PCC 7001 producing α-carboxysomes. (C) Thin section of cyanobacterium *Synechococcus* PCC7942 producing β-carboxysomes (Rae *et al*., 2013).

the chemoautotroph, *Halothiobacillus neapolitanus*, the component for α-carboxysome are encoded in a Cso cluster, where the structural proteins are located downstream of the RuBisCO large (cbbL) and small (cbbS) subunit genes (Cannon *et al.*, 2001). Typically β-carboxysomes are larger than the α-carboxysomes, which have diameters of 200 – 400 nm and 100 – 134 nm respectively (Ludwig *et al*., 2000; Rae *et al.*, 2012; Cai *et al.*, 2014).

The assembly pathway for the carboxysomes has been characterized for the β-carboxysome, which forms from the inside out. It is started when RuBisCo forms an aggregate with an encapsulation peptide of a proteins-enzyme complex. The encapsulation peptide is composed of a short C-terminal extension of about ~17 amino acids. This peptide sequence is predicted to form an amphipathic α-helix. Encapsulation peptides interact with shell proteins for the assembly of the carboxysome (Aussignargues *et al.*, 2015). However, deletion of RuBisCo does not prevent the assembly of carboxysomes (Menon *et al.*, 2008). This suggests that different interactions may play significant roles in the assembly of the microcompartment. Recently, new investigations have focused on studying the interactions between RuBisCO subunits and shell/scaffold proteins of the α-carboxysome from *H. Neapolitan* and revealed two type of protein-protein interactions in the process of α-carboxysome assembly (He *et al.*, 2018). One interaction is between the small subunit (CbbS) of RuBisCo and the scaffold protein CsoS2. The other interaction is between the N-terminal of the large subunit (CbbL) of RuBisCo and all major shell proteins CsoS1A, 1B and 1C. Factors affecting BMC formation could also include hydrophilic-hydrophobic interactions, as well as the potential energy surface of the proteins, which

underlies most calculations of structure, dynamics and thermodynamics in molecular science (Wales, Clary and Schön, 2005). These observable interactions are the driving force behind micelle formation and virus capsid assembly (Fejer *et al.*, 2009).

Proteins are complex molecules whose structures reflect the various types of chemical bonds that mediate their intricate topological arrangements. Atoms are the smallest building blocks of matter that constitute the chemical elements. Atoms bind together through different types of chemicals bonds to generate molecules. The major chemical interactions are facilitated by ionic bonds, covalent bonds and hydrogen bonds. An ionic bond is formed when an atom gains or loses one or more of its valence electrons to another atom. Thus both gain a charge and become attracted to each other (the key is the attractive force between the negatively changed ion and the positively charged that formed the bond). A covalent bond is formed when atoms share valence electrons and, as atoms do not always share electrons equally, two types of covalent bonds have emerged. One is termed a non-polar covalent bond and is formed when electrons are shared by two atoms equally, e.g. in hydrogen gas, which has two atoms that share electrons equally because they are identical and gain similar polarity. The peptide bond is a non-polar covalent bond that plays an essential role in biology. The second type is termed a polar covalent bond, which occurs when electrons are not equally shared between two atoms, e.g. in hydrogen fluoride (HF), which forms a polar covalent bond between the hydrogen atom and the fluorine atom, which is the more electronegative atom. Thus the side of the molecule where the hydrogen atom is located is more positive whereas the fluorine atom,

because it is more closely associated with the electrons, is more negative. Finally, a hydrogen bond, which is the third major type of chemical bond, is formed when a hydrogen atom presents a slight positive charge and the other atoms in the same molecule or a different molecule present a slight negative charge. These chemical bonds are very important in biology and are responsible for protein structure. For example, the peptide bond (non-polar covalent bond) forms the primary structure of the protein by joining the amino acids (building blocks of protein) together in a chain of varying length depending on the protein type. Moreover, hydrogen bonds give rise to the secondary structure (α-helices and β-sheets) in proteins. Ionic bonds allow electrostatic interactions between the side chains of the various amino acid residues either within the polypeptide backbone or between different polypeptide chains.

1.2.1 Carboxysome shell proteins

Hexameric, pentameric, and trimeric (pseudohexameric) shell proteins are the key building blocks for all BMCs including the carboxysome. *H. neapolitanus* form α-carboxysomes from six shell proteins called CsoS1A, CsoS1B, CsoS1C, all of which are BMC-H, CsoS4A, CsoS4B, both of which are BMC-P, and CsoS1D, which is a BMC-T. *Synechocystis sp* bacteria form β-carboxysomes from six shell proteins named Ccmk1, Ccmk2, Ccmk3, Ccmk4, all of which are BMC-H, CcmL, which is a BMC-P, and CcmO, which is a BMC-T (Kinney *et al*., 2011). Sequence analysis showed significant homology in the gene products of CsoS1, CcmK and CcmO (Cannon *et al.*, 2001).

1.2.1.1 CcmK

Ccmk1, Ccmk2, Ccmk3 and Ccmk4, are relatively small proteins and are believed to be the major constituents of the flat facets of β-carboxysomes. The first shell proteins structures resolved from the carboxysome were Ccmk2 and Ccmk4 from the β-carboxysome of the β-cyanobacteria (a division of the cyanobacteria defined to contain a form IB of RuBisCO) *Synechocystis* sp. PCC6803, which also forms Ccmk1 as well as Ccmk3 (Kerfeld *et al.*, 2005). These four proteins all show more than 85 % sequence similarity. The CcmK proteins are highly similar to the PduA protein of *S. enterica* propanediol utilisation BMC (Cannon *et al.*, 2001). Crystal and TEM structures analysis, of CcmK protein indicates that this protein is a single-domain and forms bidimensional (2D) layers of packed hexameric proteins containing a central pore (Kerfeld *et al.*, 2005; Truan *et al.*, 2017). In addition, the three-dimensional crystal structures analysis has revealed that the Ccmk2 hexamer units assemble together in uniform orientations to form an almost solid molecular sheet, with all of the convex faces pointing to same direction, while the Ccmk4 hexamer units form linear strips of alternating orientation between convex and concave faces **(Figure 1.5)** (Kerfeld *et al.*, 2005). Usually positively charged either arginine or lysine residues are found in the central pore of the shell protein, which suggests that these central pores play a role in the transport and sequestering of metabolites and intermediates (Kerfeld *et al.*, 2005). The central pore is ~ 4 – 7 Å in diameter and appears to be suitable for negatively charged molecules such as ribulose bisphosphate, phosphoglycerate, and HCO_3^-, but unsuitable for carbon dioxide as well as oxygen which are uncharged molecules.

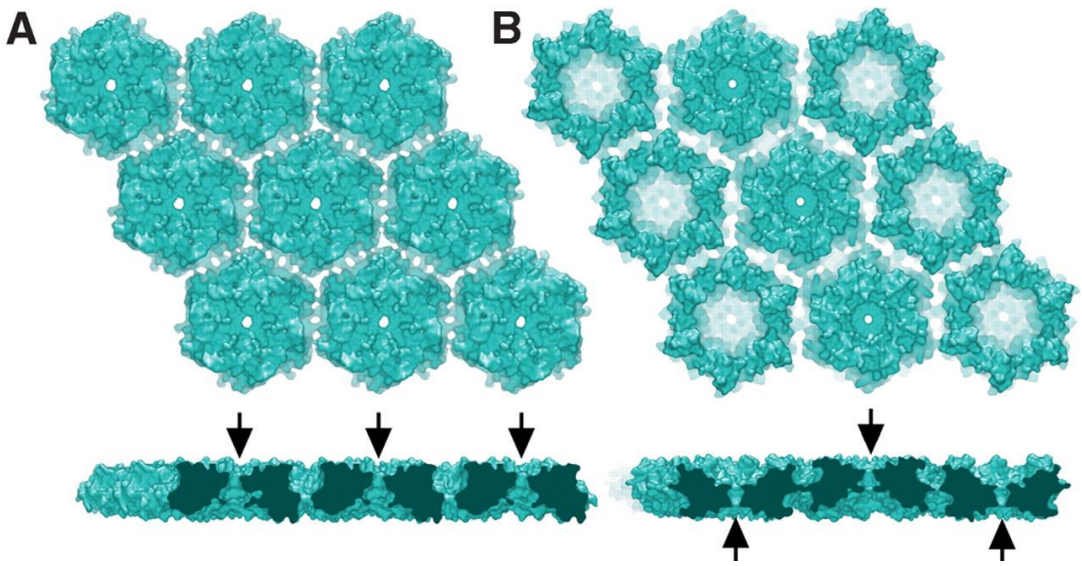


Figure 1.5: Crystal packing of Ccmk shell proteins in molecular layers. (A) CcmK2 hexamers packed in constant orientations and black arrows pointing at the convex faces. (B) CcmK4 hexamers in one of the two crystals packed in strips of alternating orientation. In the side view of each sheet. All the black arrows pointing at the pore position. This diagram was adapted from (Kerfeld *et al*., 2005).

1.3 Metabolosomes

Initially it was thought that the carboxysome was the only bacterial microcompartment. However, many decades after the discovery and characterisation of the carboxysome, different compartmentalised metabolic processes were discovered in *S. enterica* for 1,2-propanediol and ethanolamine degradation as well as in *Clostridium phytofermentans* for L-fucose and L-rhamnose utilisation (Chen *et al*., 1994; Stojiljkovic *et al*., 1995; Petit *et al.*, 2013). Investigations revealed that these structures were only formed in specific media and they were therefore often missed during earlier studies on these organisms. Further investigations revealed that these BMCs are far more complex than the carboxysome. The organelles were named “metabolosomes” as they compartmentalise a specific metabolic process that

includes several functional enzymes. Genetic studies indicated that some of the metabolosome genes are homologous to the carboxysome shell protein genes (Chen *et al.*, 1994). Indeed, carboxysome shell proteins were found to have more than 50 % similarity to metabolosome shell proteins genes. In this section three types of metabolosomes will be discussed: the functionally well characterised ethanolamine utilisation (Eut) system, the 1,2-propanediol utilisation (Pdu) system as well as the less-well characterised Fucose/ rhamnose utilisation (Cphy) system, which is the focus of this study.

1.3.1 Propanediol utilisation

For decades the carboxysome was thought to be the only specialised organelle in bacteria used to increase the efficiency of cell growth. As discussed previously, sequence analysis revealed proteins that were highly homologous to carboxysome shell proteins such as PduA, which was found to be encoded within an operon involved in 1,2-propandiol degradation in *Salmonella typhimurium*. This protein has 55 % amino acid identity and 70 % similarity to the CcmK shell protein of the carboxysome from *Synechococcus sp*. (Chen *et al*., 1994). Further studies revealed that when *Salmonella* was grown in the presence of 1,2-PD structures of between 100 to 150 nm in diameter that resembled carboxysomes were observed **(Figure 1.6)** (Chowdhury *et al*., 2014). The facets of the Pdu metabolosome are thought to consists of hexameric, trimeric and pentameric shell proteins (Chowdhury *et al*., 2014). These metabolosomes sequester metabolic enzymes, metabolites and cofactors required for 1,2-PD degradation with an overall mass of approximately 600 MDa (Cheng *et al.*, 2008). Studies suggested

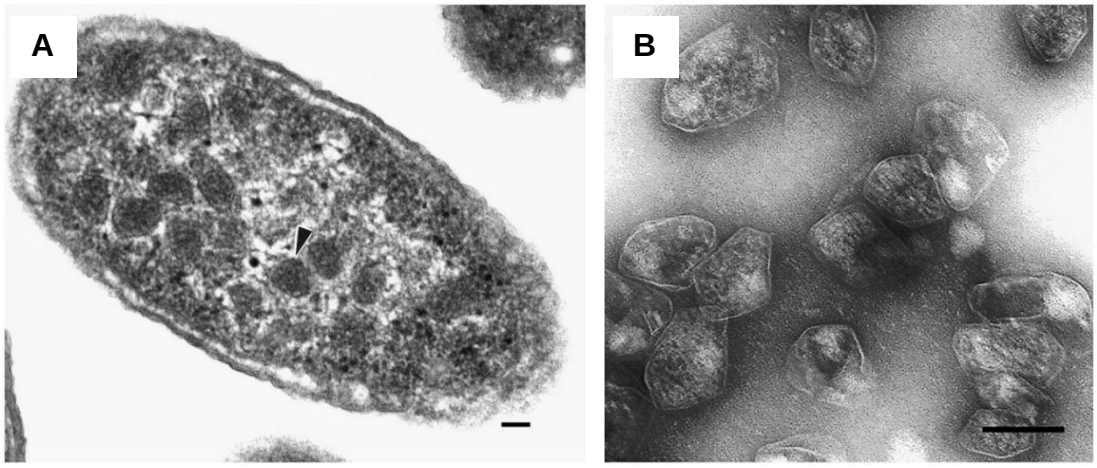


Figure 1.6: TEM analysis of the Pdu microcompartment. (A) Thin section of *S. enterica* cell producing BMCs with an arrow pointing at one of the BMCs. (B) Purified BMCs isolated from *S. entrica* (Scale bar: 100 nm) (Crowley *et al*., 2008).

that the main function of the Pdu metabolosome is to optimise growth of the bacterium on 1,2-PD via encapsulating a toxic pathway intermediate (propionaldehyde), thereby avoiding carbon loss and DNA damage (Bobik *et al*., 2015). Under standard culture conditions, high-pressure liquid chromatography (HPLC) analysis and growth tests showed that propionaldehyde accumulates to toxic levels in BMC-minus mutants of *S. enterica* during growth on 1,2-propendiol. In these mutants only a small fraction (13 %) of the propionaldehyde (boiling point of 48°C) was lost due to volatility (Sampson and Bobik, 2008). The BMC does not appear, therefore, to have a major effect on retaining the volatile intermediate within the confines of the shell. Hence, this suggests that the main function of the Pdu metabolosomes is to protect the cell from the toxicity of the propionaldehyde, which can modify both protein and DNA through its chemical reactivity.

Metabolosomes are also though to enhance the gut invading pathogen proliferation through the coordinated action of type III secretion (~ 10 % – 35

% of cells) and BMCs formation (~ 65 % – 90 % of cells) (Jakobson *et al*., 2016). Each pathogenic cell undergoes a fate decision between type III secretion and BMCs formation. Invading cells adjust the chemical environment of the intestine to become more favourable to *Salmonella* proliferation by inducing gut inflammation, which causes thiosulfate oxidation to tetrathionate. Moreover, tetrathionate is the alternative electron acceptor that can support the anaerobic growth on 1,2-PD of *Salmonella* via using endogenously synthesized B_{12} (Price-Carter *et al.*, 2001).

Fucose and rhamnose are naturally occurring deoxy sugars found in the glycans of plant, mammalian and insect cell walls (Petit *et al.*, 2013). The major fermentation products of these common plant sugars is 1,2-PD, which can act as a carbon source in anaerobic environments such as those found in the intestines of higher animals and sediments,(Obradors *et al.*, 1988). The 1,2-propandiol utilisation operon is not only found in *S. enterica* but also present in multiple enteric and soil-dwelling species of bacteria such as *Citrobacter*, *Klebsiella*, *Shigella*, *Yersinia*, *Listeria*, *Lactobacillus*, *Lactococcus*, *Escherichia coli* (E24377A) as well as *Clostridium* (Bobik *et al*., 2015). It is the *C. phytofermentans* fucose/rhamnose operon that is the focus of this study.

In *S. enterica* the Pdu locus includes 24 genes important for BMC assembly and 1,2-PD degradation (Chowdhury *et al*., 2014). Eight of these genes encode for 9 BMC shell proteins (PduA, B, B′, J, K, M, N, T and U). PduV is thought to play a role in the spatial organisation of the BMC (Parsons *et al.*, 2010). Seven of the genes encode for 1,2-PD utilisation enzymes (PduC, D, E, L, P, Q and W), while a further five encode for enzymes associated with

coenzyme B_{12} (Ado-B_{12}) assimilation and recycling (PduG, H, O, S and X). Two genes encode for a transcriptional regulator and 1,2-PD diffusion facilitator (PocR and PduF respectively) **(Figure 1.7)** (Chowdhury *et al.*, 2014; Sturms *et al.*, 2015).

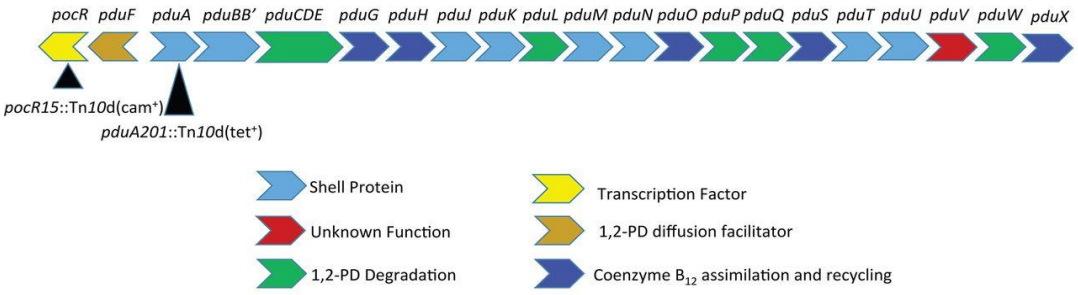


Figure 1.7: Schematic of the 24 gene pdu operon of *Salmonella* responsible for BMC assembly and 1,2-PD degradation (Sturms *et al.* 2015).

The Pdu BMC is thought to operate by allowing 1,2-PD to enter the lumen of the metabolosome through a selectively permeable pore in the PduA shell protein, thereby allowing access to the substrate and preventing the exit of propionaldehyde (Sawaya *et al.*, 2015). 1,2-PD is then converted into the toxic intermediate propionaldehyde by the coenzyme B_{12}-dependant diol dehydratase (PduCDE) **(Figure 1.8)** (JETER, 1990; Roth *et al.*, 2016).

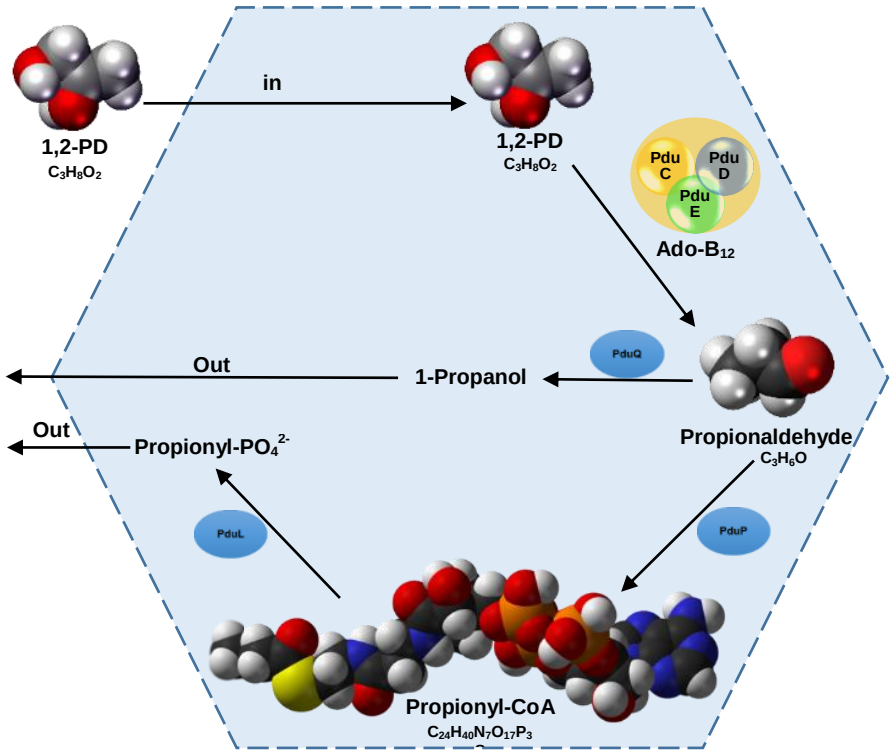


Figure 1.8: Model for the Pdu metabolosome where 1,2-PD degradation pathway occurs inside the lumen to optimise growth and confining volatile metabolic intermediates (propionaldehyde).

Subsequently, propionaldehyde is converted to propionyl-CoA and 1-propanol via an aldehyde dehydrogenase (PduP) and 1-propanol dehydrogenase (PduQ) respectively. During fermentation, 1-propanol is excreted in order to balance excess reducing equivalents that are generated via the conversion from propionaldehyde (Price-Carter *et al.*, 2001). A phosphotransacylase (PduL) catalyses the conversion of propionyl-CoA to propionyl-phosphate (propionyl-PO_4), which is then converted to propionate and ATP via propionate kinase (PduW) (Palacios *et al.*, 2003; Liu *et al.*, 2007). Under fermentation conditions, *S. enterica* can't use 1,2-PD as its sole carbon and energy source because of 1-propanol excretion (Price-Carter *et al.*, 2001). However, under respiratory growth of either aerobic or

during anaerobic with tetrathionate as the terminal electron acceptor, *S. enterica* is able to use 1,2-PD as its carbon and energy source (JETER, 1990; Price-Carter *et al.*, 2001).

The assembly pathway for the Pdu microcompartment is thought to involve encapsulation peptides which are often present as N-terminal extensions of about ~ 18 amino acid found on some of the cargo enzymes such as PduP that are found within the BMC. These α-helical encapsulation peptides are separated from enzymatic domains via a poorly conserved linker region and are sufficient to encapsulate proteins within microcompartments (Aussignargues *et al.*, 2015).

1.3.1.1 Reactivation and recycling of diol dehydratase

Adenosylcobalamin (cobalamin, coenzyme B_{12}, and Ado-B_{12}) is an essential component for 1,2-PD degradation. Some bacteria have the ability to synthesize cobalamin and use it in their metabolic processes such as enteric bacteria. For 1,2-PD degradation, *S. enterica* can synthesize, reactivate, use, and recycle Ado-B_{12} anaerobically (Roth *at al*., 1996; Chowdhury *et al*., 2014). However, in the presence of oxygen *S. enterica* can only use Ado-B_{12} but not synthesize it for either 1,2-PD or ethanolamine degradation (Roth *et al*., 1996).

1,2-PD is an essential supplement for Ado-B_{12} synthesis, because it induces the *cob* operon (Ado-B_{12} synthesis genes) (Roth *et al*., 1996; Price-Carter *et al.*, 2001). Ado-B_{12} has a complex structure with a molecular mass of 1580 Da. The nutrient can undergo both recycling and reactivation within the BMC

(Roth *et al*., 1996; Chowdhury *et al*., 2014). These processes happen inside the lumen of the BMC during diol dehydratase catalysis, which is the first enzyme in 1,2-PD utilisation. The upper adenosyl ligand is sometimes lost during the catalytic cycle and is replaced with a hydroxyl group (Chowdhury *et al*., 2014). Subsequently, PduGH reactivatase converts the inactive form OH-B_{12} diol dehydrase to the active form of Ado-B_{12} by removing the OH-B_{12} from the diol dehydrase. Two more enzymes that are essential to the B_{12} reactivation include PduO (adenosyltransferase) and PduS (cobalamin reductase), which are responsible for the transformation of OH-B_{12} into Ado-B_{12} (Sampson *et al*., 2005).

1.3.1.2 Pdu shell proteins

In this section the shell proteins of metabolosomes that have sequence similarity to the shell proteins of *C. phytofermentans* (PduU, A, K, N, and T) will be discussed.

The *S. enterica* BMC shell protein **PduU** has 45 % identity to Cphy1176 of *C. phytofermentans* **(Section 3.2.3.1)**. This protein is composed of 116-amino-acids, and houses a single-BMC-domain. PduU was reported as one of the less abundant proteins in the *S. enterica* Pdu BMC system (Havemann and Bobik, 2003). However, it was the first metabolosome shell protein to be crystallised (Crowley *et al*., 2008). The overall structure is similar to the hexamers formed by CcmK1-4. However, its 3D structure revealed a circularly permuted BMC fold and it contains a unique six stranded beta barrel over the central pore (Crowley *et al*., 2008; Laganowsky *et al*., 2012).

This unique structure suggests that PduU may serve a particular role within the BMC rather than a role in transport (Crowley *et al.*, 2008).

The *S. enterica* **PduA** protein shares 42 % and 90 % identity with the (*C. phytofermentans*) Cphy_1180 and Cphy_1182, respectively, **(Section 3.2.3.2)**. PduA is one of the major Pdu shell proteins and is 94-amino-acids in length. Crystal structure analysis reveals a symmetric homohexamer, similar again to the carboxysome shell proteins CcmK1-4, with a central pore that is 5.6 Å in diameter. The central pore is assumed to facilitate transport of 1,2-PD into the lumen of the BMC (Yeates *et al.*, 2010). Physiological and electron microscopy studies on *S. enterica* using mutants lacking PduA have demonstrated that the absence of this shell protein does not prevent the assembly of BMCs as previously thought, but formed enlarged BMCs with disrupted integrity (Cheng *et al.*, 2011). When PduA is expressed without the other Pdu shell proteins it forms nanotubes that are approximately 20 nm in diameter **(Figure 1.9)**.

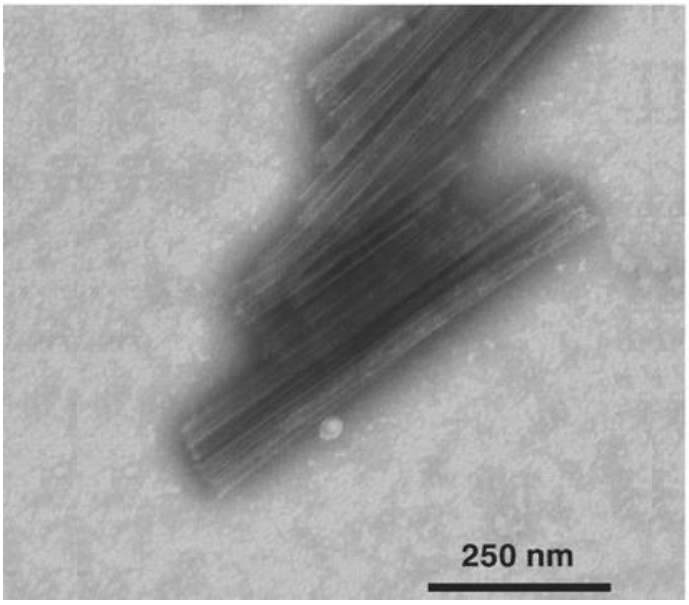


Figure 1.9: TEM image of purified *C. freundii* PduA shell protein recombinantly expressed in *E. coli* (Uddin *et al*., 2018).

Three models have been proposed to produce a nanotube consistent with the experimentally determined dimeter of 20 nm. The first model is helical in nature and is formed from 10 PduA hexamers per turn, with a bend angle of 37.5° and a pitch of 61 nm **(Figure 1.10)**. In this structure the concave surface of the hexamers faces out. The other model is the zigzag with 10 PduA hexamers per turn having 36° bend angle. The last model is the armchair with 12 hexamers per turn having a bend angle of 30° *(*Uddin *et al*., 2018).

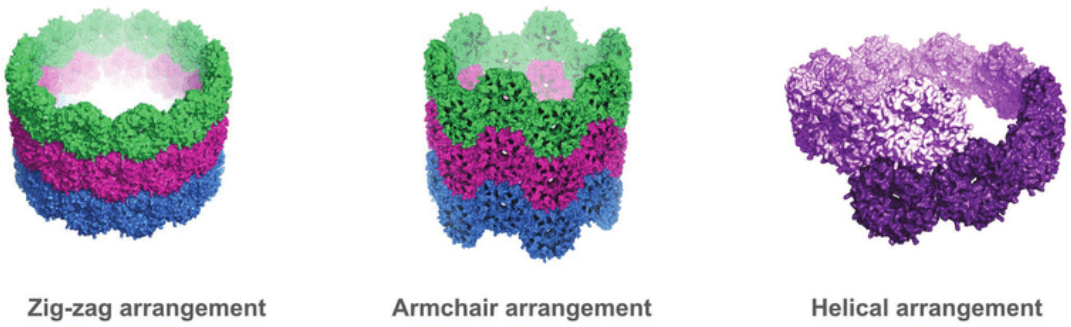


Figure 1.10: The three suggested models for PduA-derived nanotubes (Uddin *et al*., 2018).

PduK is a minor Pdu shell protein and it forms a hexamer from six single BMC domain subunits. It is composed from 160 amino acids and in comparison to other BMC-H proteins it contains a C-terminal extension (~ 70 amino acids) (Chowdhury *et al.*, 2014). Deletion mutants of PduK did not substantially affect BMC assembly or propionaldehyde accumulation, but its absence caused BMC aggregation (Cheng *et al.*, 2011) Sequence analysis shows a potential Fe-S binding motif (CNLCLDPKCPRQKGEPRTLC) in the

C-terminal extension and experimental evidence shows that the purified PduK was able to bind to iron. However no firm evidence for an Fe-S cluster has been presented (Crowley *et al.*, 2010).

PduN is a pentamer that forms the vertices of the BMC and is a small component (2 %) of the BMC proteome (Mayer *et al.*, 2016). The deletion of PduN causes grossly abnormal BMC structures and impacts growth on 1,2-PD (Cheng *et al.*, 2011). It is a pentameric BMC shell protein similar to CcmL (β-carboxysome), CosS4 (α-carboxysome), and EutN (ethanolamine) (Sutter *et al.*, 2013).

PduT is a Pdu pseudohexamer shell protein that belongs to the BMC-T class that has a double-BMC-domain similar to PduB. It is composed of 184 amino acids but represents only 3 % of the BMC proteome (Chowdhury *et al*., 2014). It is therefore a minor Pdu shell protein and its deletion does not impact upon the BMC structure. Crystal structure analysis revealed that this hexagonal shape protein has a central pore suitable for binding a [4Fe-4S] cluster (Crowley *et al.*, 2010; Pang *et al*., 2011). Sequence analysis showed four cysteine residues (C_{38}, C_{108}, C_{110}, and C_{136}), but only C_{38} is responsible for the metal cluster by providing three ligands from the three protein subunits (Pang *et al*., 2011).

As mentioned previously, shell proteins that are located in the cluster of fucose/rhamnose degradation in *C. phytofermentans* share similar features with other shell proteins that are located in the cluster of 1,2-Propanediol, ethanolamine utilisation also with carboxysomes **(Figure 1.11)**.

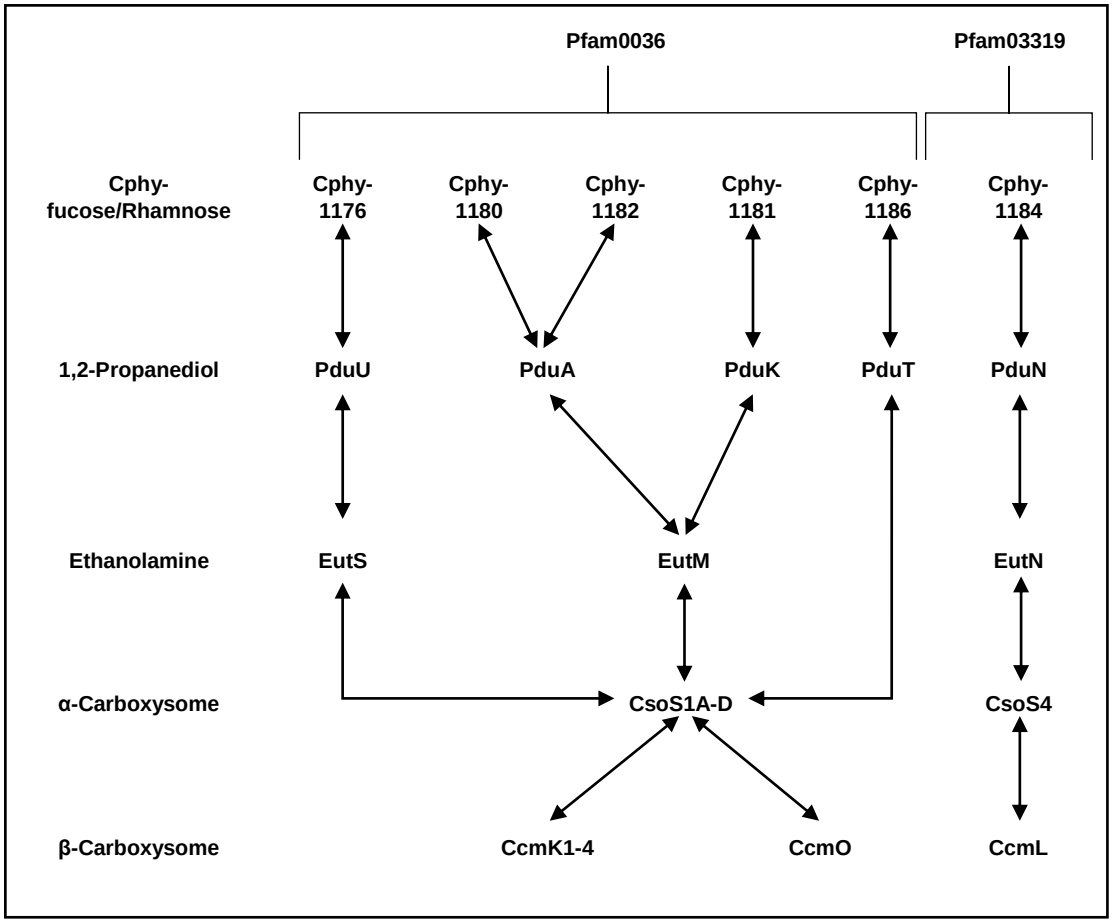


Figure 1.11: A diagram showing the relationship between shell proteins of the microcompartments (Frank *et al.*, 2013; Petit, *et al.*, 2013).

1.4 Ethanolamine utilisation

Not many studies been conducted in the area of ethanolamine metabolosome biology in comparison to the work that has been reported as Pdu metabolosomes and carboxysomes. Ethanolamine is an important organic chemical compound for many bacteria that live in the gastrointestinal tract of mammals as a source of carbon, nitrogen, and energy. Ethanolamine

is produced from the degradation of the membrane component phosphatidyl ethanolamine, which is a class of phospholipids (Kerfeld *et al*., 2010).

TEM analysis revealed polyhedral structures in thin-sections of *S. enterica* (strain JE8392 *eut*$^{+}$) grown on ethanolamine as a source of carbon and energy **(Figure 1.12)** (Brinsmade *et al*., *2005*). Sequence and mutational analysis on Eut (ethanolamine utilization) operon has revealed a significant degree of similarity to several proteins involved in the formation of

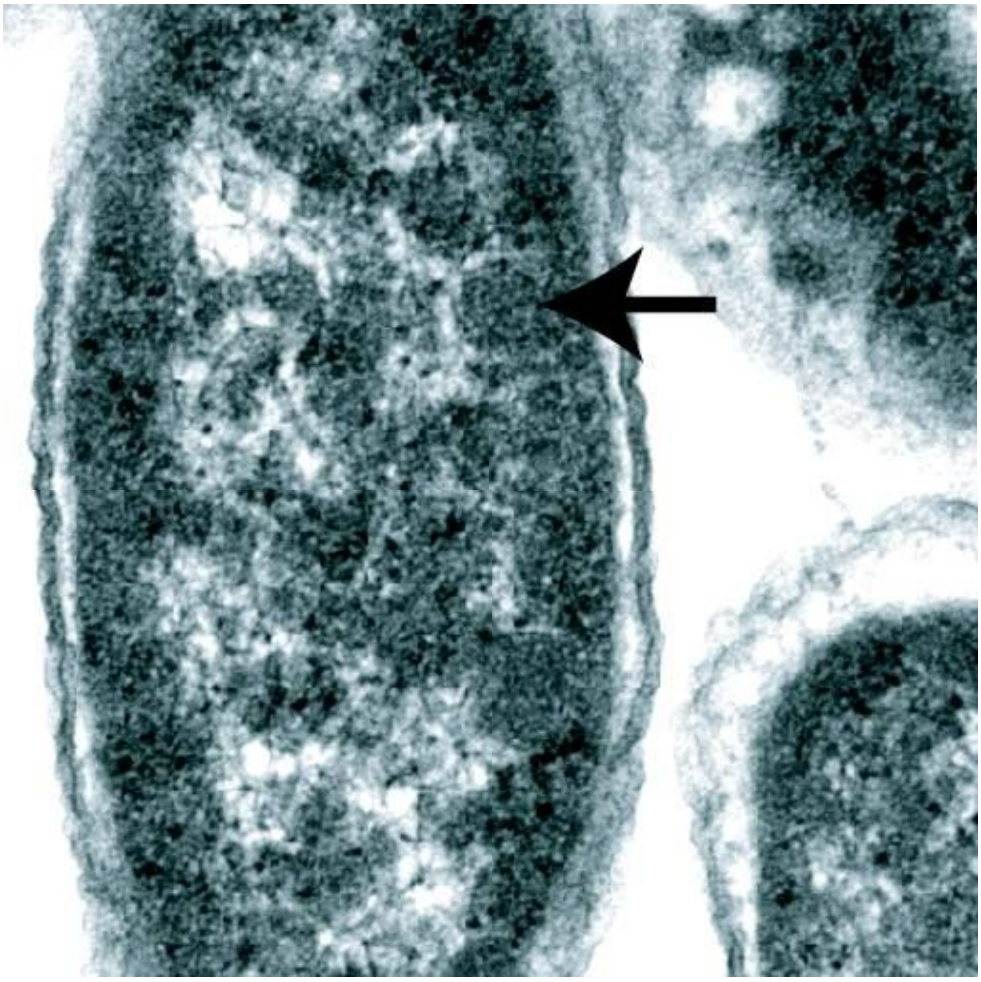


Figure 1.12: TEM analysis of of *S. enterica* (strain JE8392 eut^{+}) grown in 30 mM ethanolamine. Arrow is pointing at one of the BMC structures (Brinsmade *et al*., 2005).

carboxysomes or the 1,2-propanediol utilisation operon (Stojiljkovic *et al*., 1995). **Figure 1.13** shows the Eut operon in *S. enterica* that houses 17 genes important for BMC assembly and ethanolamine degradation (Kofoid *et al.*, 1999). Under standard conditions (37°C; pH 7.0) six of the genes (EutA, B, C, D, E, and R) can form the BMCs with the remaining genes (EutM, N, L, K, and G) becoming more important under certain conditions such as higher pH and rapid gas exchange (Penrod and Roth, 2006).

To some extent there are parallels between the chemistry of propanediol and

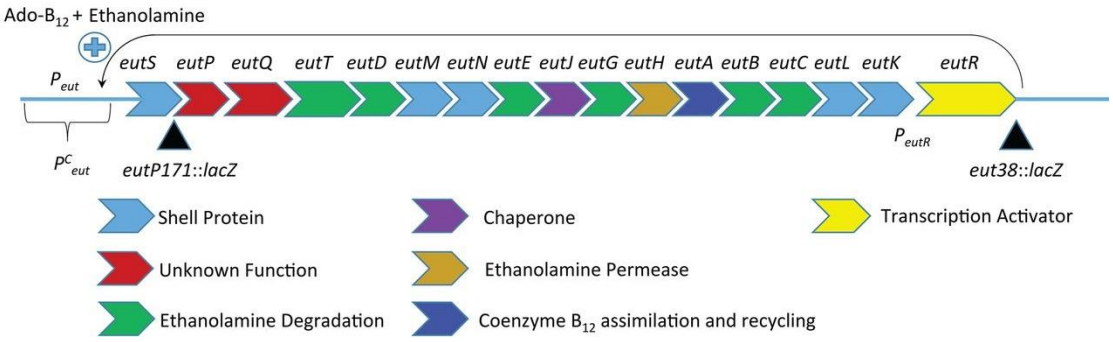


Figure 1.13: Schematic of the 17 gene *eut* operon of *Salmonella* responsible for BMC assembly and ethanolamine degradation (Sturms *et al.* 2015).

ethanolamine utilisation. Ethanolamine enters the microcompartment after it is has been transported into the cytoplasm by the ethanolamine facilitator EutH (Tsoy *et al*., 2009). Ethanolamine is then converted into acetaldehyde and ammonia by the action of the EutBC complex, which is a B_{12}-dependant enzyme. EutB is the ethanolamine ammonia-lyase large subunit whilst EutC is the ethanolamine ammonia-lyase small subunit **(Figure 1.14)** (Mori *et al.*, 2004). Subsequently, the acetaldehyde is transformed into ethanol and

acetyl-CoA via EutE (aldehyde dehydrogenase) and EutG (Kerfeld *et al*., 2010). As with the Pdu metabolosome, it is suggested that the Eut metabolosome is able to protect the cells from toxicity of acetaldehyde (Penrod and Roth, 2006).

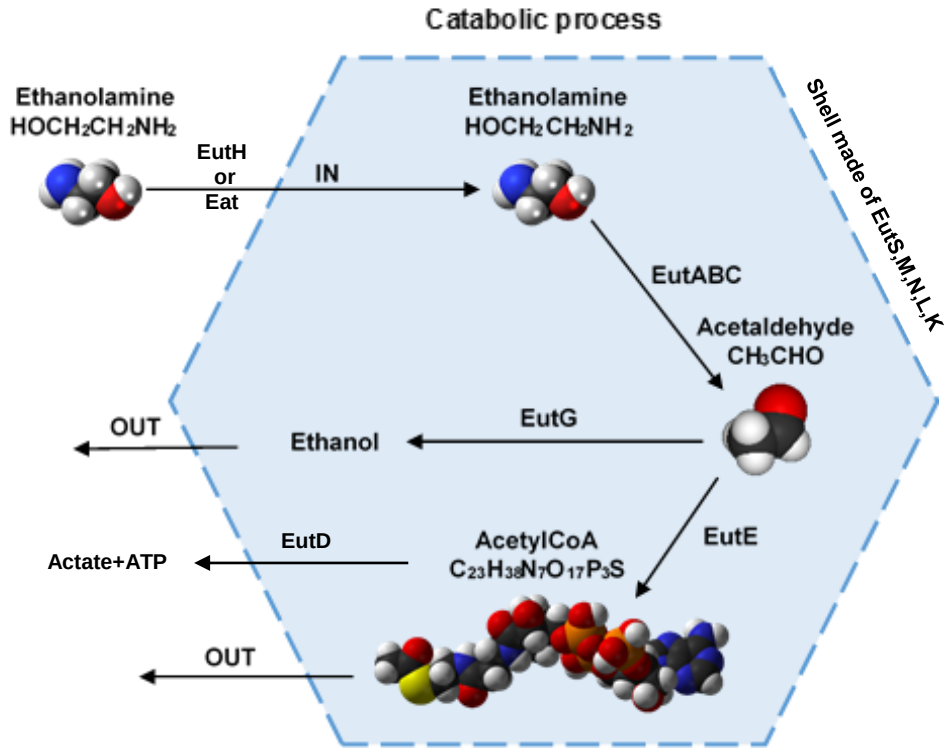


Figure 1.14: A model for the Eut metabolosome where ethanolamine degradation happens to optimise growth and confining volatile metabolic intermediates (acetaldehyde).

Based on sequence analysis, both the Pdu and Eut operons contain genes encoding BMC-domain proteins and enzymes for a degradative metabolic pathway that proceeds through an aldehyde intermediate (propionaldehyde and acetaldehyde) (Cheng *et al.*, 2008). Subsequently, it was suggested that the major role of metabolosomes might be to form compartments that

sequester aldehydes, which were known to be toxic component and volatile. The boiling point of propionaldehyde is 48°C which, is less volatile than acetaldehyde (acetaldehyde has a boiling point of 20°C). However, mutations in the ethanolamine utilisation system that lack BMCs proteins has shown that acetaldehyde lost during growth has no obvious toxic effect, which has led to the suggestion that the main function of the Eut metabolosome is to minimize acetaldehyde loss and thereby improve growth efficiency (Penrod and Roth, 2006).

1.5 Fucose and rhamnose utilisation by *C. phytofermentans*

Fucose and rhamnose are described as sugars earlier in the introduction. Although, most sugars in nature are found in the D- enantiomer, form fucose and rhamnose are found in nature in the L- enantiomer form. When grown on fucose/rhamnose, the fermentation pathway of *C. phytofermentans* produces a variety of products including ethanol, acetate, lactate, formate, propionate, and propanol. *C. phytofermentans* strain ISDg **(Figure 1.15)** is an obligate anaerobic, mesophilic, cellulolytic bacterium, which was initially isolated from forest soil near the Quabbin Reservoir in Massachusetts (USA) (Warnick *et al*., 2002). It exists as a long, thin, straight, motile rod-shaped Gram-positive bacterium. Many sugars can support its growth, including arabinose, glucose, and cellulose (Warnick *et al*., 2002). As *C. phytofermentans* has the ability to convert biomass into biofuel, it is currently being developed and used commercially to increase ethanol yield (Svetlitchnyi *et al.*, 2013).

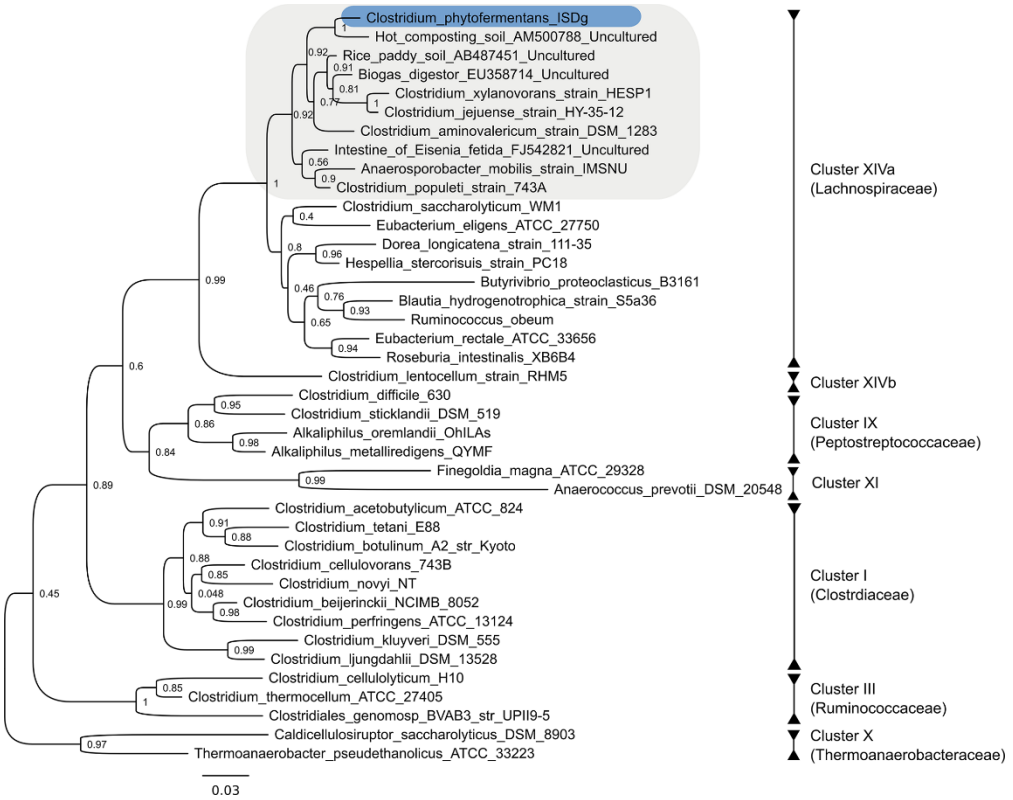


Figure 1.15: Neighbor-joining tree of *C. phytofermentans* and related taxa within the class Clostridia based on 16S rRNA gene sequences. Strain ISDg is shaded with blue in XIVa cluster (Latouf *et al.*, 2015).

In addition, TEM analysis has revealed multiple carboxysome-like polyhedral structures approximately 100-150 nm in diameter that are formed when the strain is grown on fucose or rhamnose **(Figure 1.16)** (Petit *et al.*, 2013). The genome of *C. phytofermentans* has three loci that have gene clusters encoding the shell proteins of BMCs. One of these loci shares a similar (67 – 90 %) role in the degradation of fucose of *Roseburia inulinivorans* in the human gut (Scott *et al.*, 2006; Petit *et al.*, 2013). Many of the genes that encode the BMC loci of *C. phytofermentans* are similar to genes involved within the 1,2-propanediol utilisation *pdu* operon of *S. enterica* including the BMC formation (Petit *et al.*, 2013).

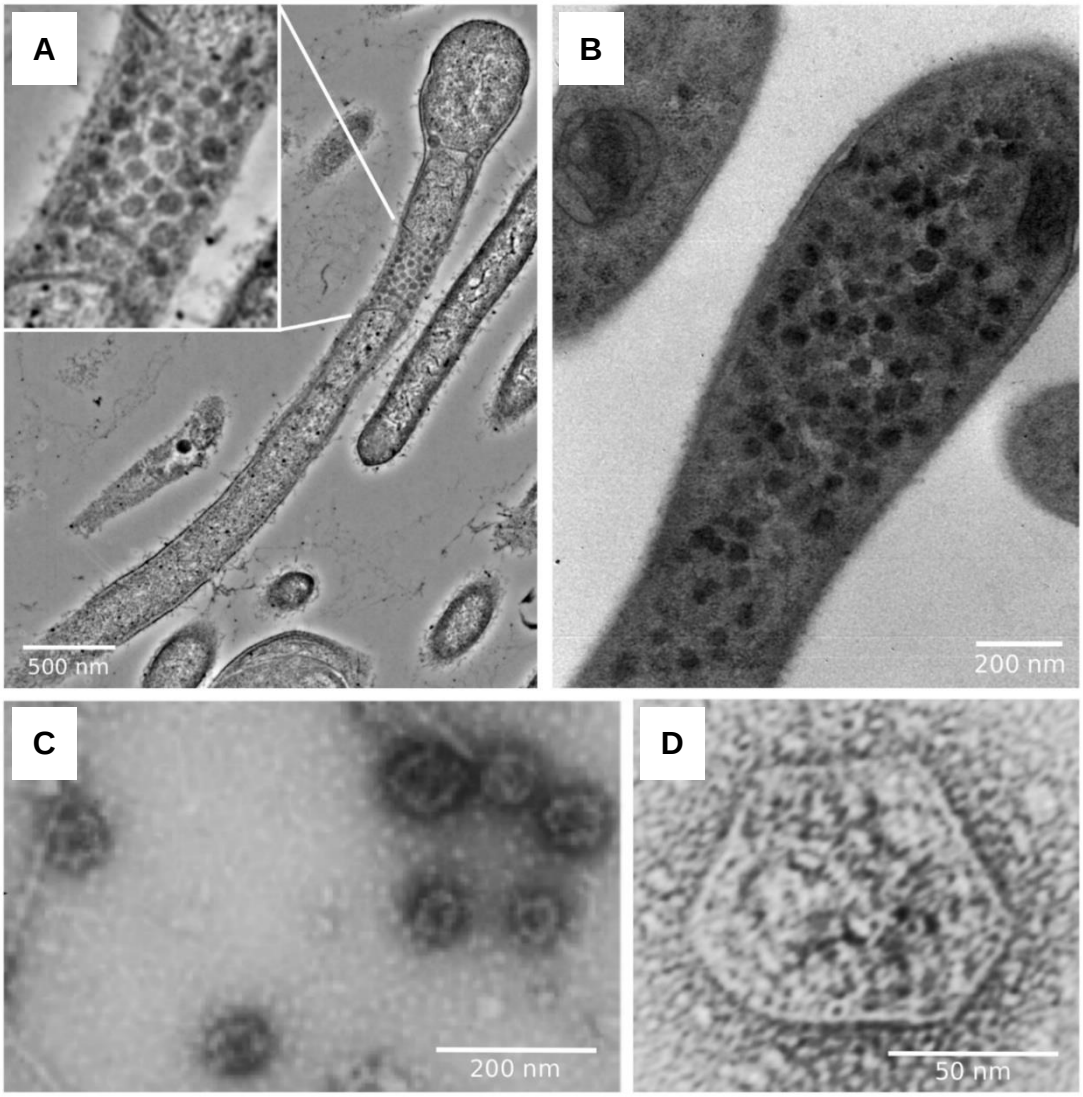


Figure 1.16: TEM analysis of *C. phytofermentans* ISDg. (A) Thin section of the bacteria grown on fucose and showing magnified BMCs. (B) Thin section of the bacteria grown on rhamnose and showing multiple BMCs. (C) negatively stained of purified BMC that looks contaminated with flagella. (D) Magnified negatively stained BMC (Petit *et al.*, 2013).

C. phytofermentans has a 14-gene operon that houses the genetic information for a fucose/rhamnose utilisation BMC. Six genes (Cphy1176, 1180, 1181, 1182, 1184, and 1186) encode for shell proteins **(Section 3.1/ Figure 3.1)**, which form the BMCs outer structure, and eight genes encode for the metabolic enzymes that are responsible for converting fucose and rhamnose into propanol, and propionate. In addition, *C. phytofermentans*

lacks the coenzyme B_{12}-dependent 1,2-propanediol dehydratase (PduCDE), as well as its reactivation cofactors (PduGH), but does contain a glycyl-radical enzyme (diol dehydratase), and its activator protein (Petit *et al.*, 2013). This is similar to the dehydratase of *Clostridium butyricum* (Scott *et al.*, 2006; Petit *et al.*, 2013). Furthermore, the *C. phytofermentans* BMC locus has a putative propanol dehydrogenase that appears to be a zinc-dependent transcriptional regulator, but they are not related to those found in the *S. enterica* Pdu locus. Instead they are related to the medium chain dehydrogenase/ reductase (MDR) and the systems of *E. coli* that usually act as repressors in sugar and nucleoside metabolism (DeoR) families respectively (Bobik *et al*., 1992; Rondon *et al*., 1995; Petit *et al.*, 2013).

1.5.1 *C. phytofermentans* fucose and rhamnose fermentation pathways

The *C. phytofermentans* pathway functions to allow the metabolism of fucose and rhamnose when they are present in the growth media. L-fucose and L-rhamnose are imported via a specific ABC transporter system of CUT2-type, which usually involved in monosaccharide transport before they are converted to fuculose- and rhamnulose-phosphate by their respective kinases, mutarotases and isomerases. Fuculose- and rhamnulose-phosphate are converted to lactaldehyde and dihydroxyacetone phosphate DHAP by Cphy1177 (Fuculose and Rhamnulose-P aldoase) **(Figure 1.17)**. DHAP is used in glycolysis and ultimately is converted into products such as ethanol, acetate, lactate, and formate. Further metabolism occurs within the BMC where lactaldehyde is converted to 1,2-PD by the enzyme Cphy1185

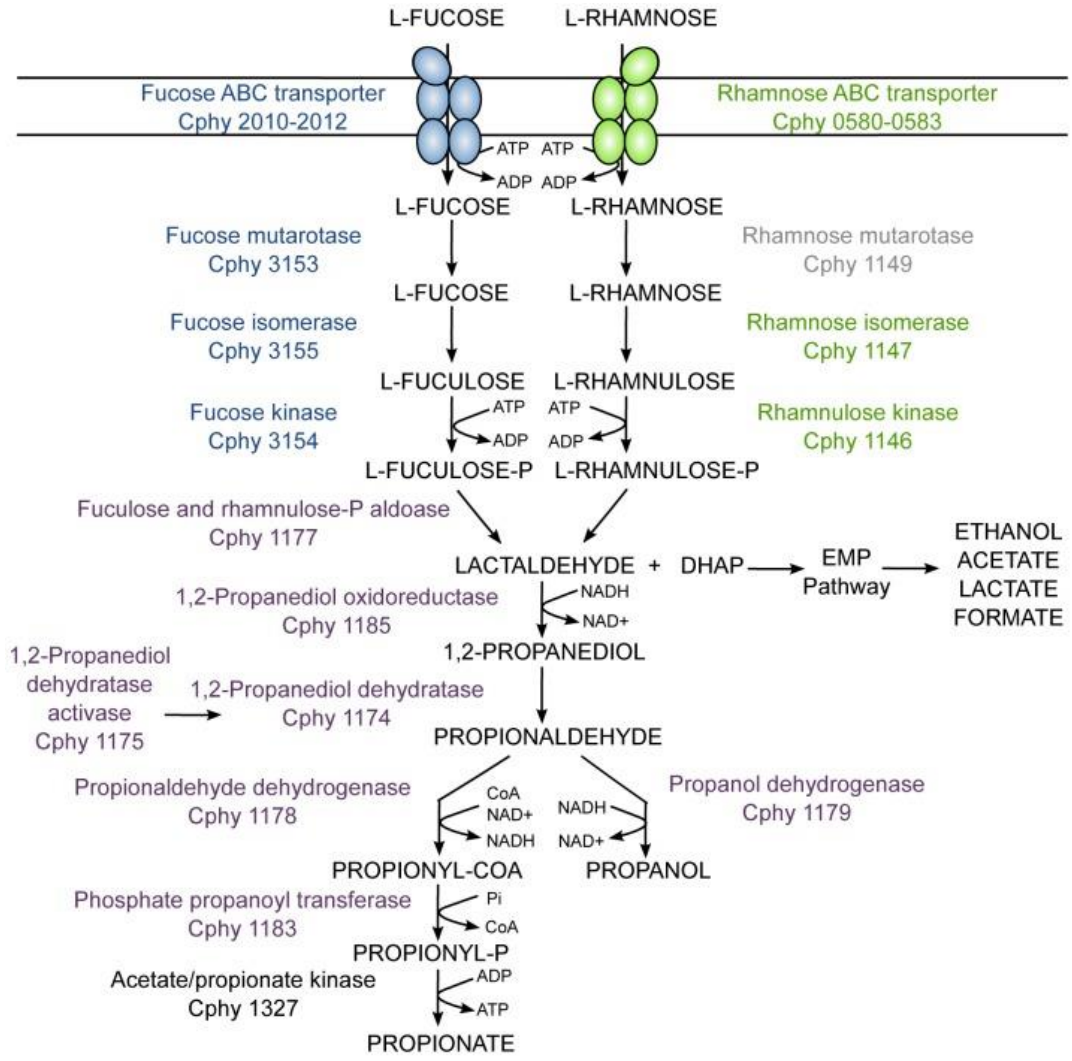


Figure 1.17: *C. phytofermentans* fucose and rhamnose fermentation pathways based on sequence homology and expression data. The genes names in purple are proposed to be internalised within a BMC. (Petit *et al*., 2013).

(PduS-like 1,2-PD oxidoreductase). Following this, 1,2-PD is transformed into propionaldehyde via the glycyl-radical enzyme Cphy1174 (1,2-PD dehydratase) and 1175 (1,2-PD dehydratase activase). Propionaldehyde can then be converted to propionyl-CoA or propanol by the enzymes Cphy1178 (propionaldehyde dehydrogenase) and Cphy1179 (propanol dehydrogenase) respectively. Finally, propionyl-CoA is transformed by Cphy1183 (phosphate

propanoyl transferase) to propionyl-P, which is then converted to propionate via Cphy1327 (acetate/ propionate kinase).

1.6 Aims and objectives

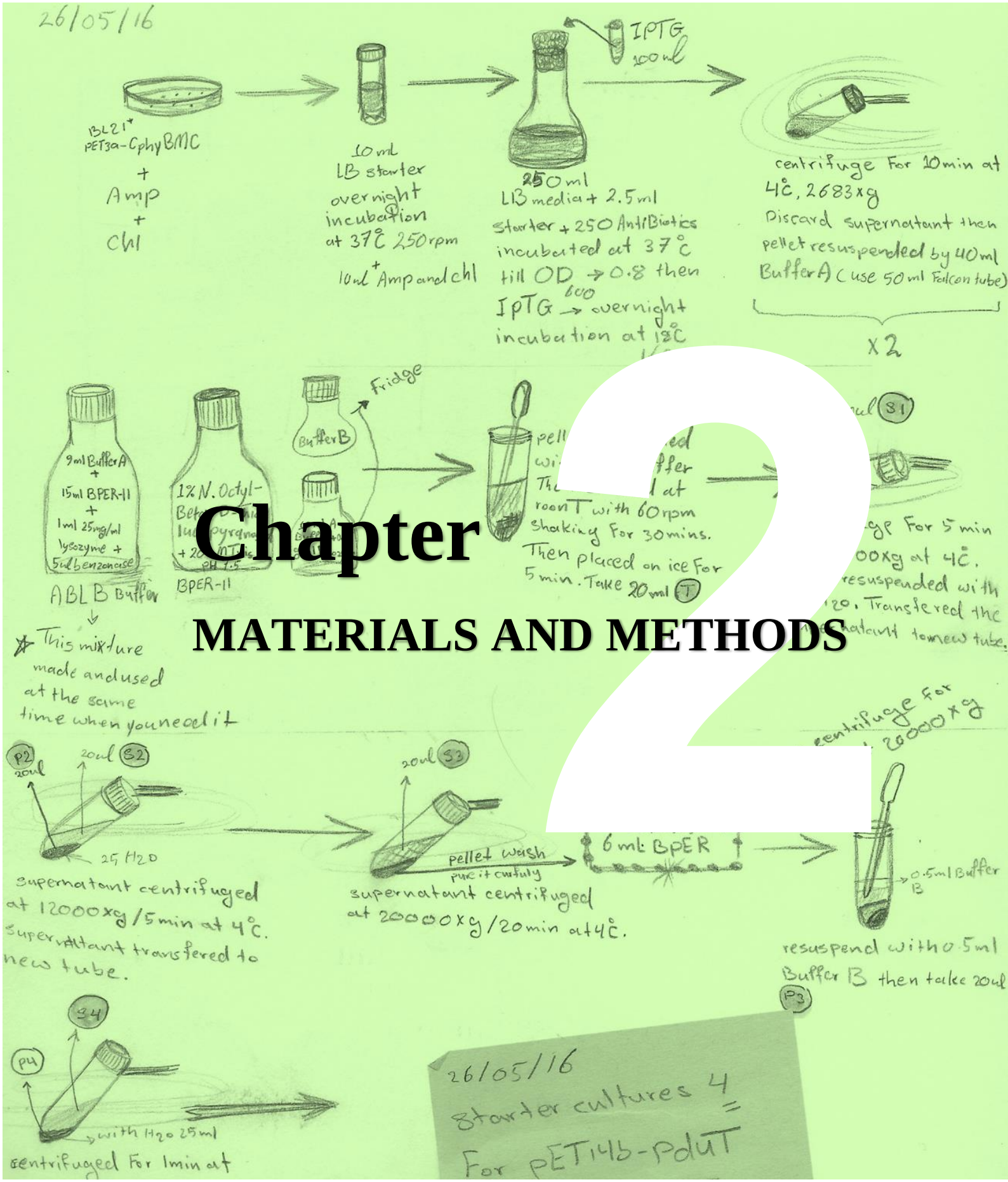
One of the key findings from the work done on *C. phytofermentans* was that the BMCs observed within the bacterium appeared to be very uniform in size and shape. This is in stark contrast to the metabolosomes observed in most other bacteria. Where BMCs have an irregular polyhedral appearance. More regular-shaped BMCs could allow for more detailed structural studies on these macromolecular organelles.

With this in mind, the overall aim of this project was to try and clone the various BMC-associated proteins of the *C. phytofermentans* fucose/rhamnose system in order to generate a recombinant BMC in *E. coli*. This was to be attempted through the two major objectives outlined below:

1) Clone and express in *E. coli* the individual shell proteins that are involved in *C. phytofermentans* BMC formation, characterise the production of the proteins and determine their ability to form functional BMCs.
2) Engineer empty BMCs from the six shell proteins and investigate the minimum number of genes required for BMC formation in *E. coli*.

Chapter 2

MATERIALS AND METHODS



2.1 Materials

2.1.1 Chemicals

The majority of chemicals and antibiotics used in this project were from Sigma-Aldrich®, OXOID Ltd, and Fisher. The rest of the materials were purchased from different suppliers such as Roche FastStart High Fidelity PCR System from Roche Diagnostics GmbH; Agarose universal from Alphalaboratories; QIAquick® Gel Extraction Kit and QIAprep® Spin Miniprep Kit from QIAGEN GmbH; restriction enzymes, T4 ligase and their buffers from Promega, except BsrGI which was purchased from New England Biolabs, Inc; cloning plasmids from Novagen; IPTG and Ampicillin from Melford Laboratories Ltd; chelating sepharose fast flow, disposable and empty PD-10 columns were purchased from GE Healthcare Life Science; crystallisation formulated screens solution for protein crystal growth structure screen 1 MD1-01 & 2 MD1-02 were purchased from Molecular Dimension.

2.1.2 Bacterial strains

All bacterial strains (or vectors) used in this study are listed in **Table 2.1**. Novagen, Promega, and Invitrogen were the main sources for the purchase of bacterial strains.

Table 2.1: List of bacterial strains used in this project.

Bacterial strain	Description and genotype	Reference / Source
BL21(DE3)	F^- *ompT hsdS$_B$ ($r_B^- m_B^-$)gal dcm (DE3)*	Studier *et al*. (1990)13 Novagen
E. coli DH5α	F^- *endA1 hsdR17 supE44 thi-1* λ-recA1 *gyrA96 relA1* Δ(*lacZYA-argF)U169* φ80 d*lacZ* Δμ15	Gibco-BRL
BL21(DE3)pLysS	F^- *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pLysS (Cam^R)	Novagen

JM109	*end*A1, *rec*A1, *gyr*A96, *thi, hsd*R17 (r_k^-, m_k^+), *rel*A1, *sup*E44, Δ(*lac-pro*AB), [F´ *tra*D36, *pro*AB, *laq*I^{q}ZΔM15]	Promega

2.1.3 Plasmids

Table 2.2: List of vectors used in this project.

Vector name	Description
pET3a	Plasmid carries an N-terminal T7-Tag sequence, Amp^R and couples of cloning sites.
pET14b	Plasmid carries an N-terminal His-Tag sequence, thrombin cleavage site, T7 expression region and couples of cloning sites.
pLysS	Plasmid carries a T7 lysozyme sequence, Cm^R and used in lDE3 lysogenic hosts to suppress basal expression from the T7 promoter.
pGEM-T-Easy	Convenient systems to clone PCR products, carries 3´-T overhangs, T7, Amp^R and SP6 RNA polymerase promoters.
pET3a-Cphy1176	PCR fragments of *Cphy1176* ligated within *NdeI* and *SpeI* of modified pET3a.
pET14b-Cphy1176	PCR fragments of *Cphy1176* ligated within *NdeI* and *SpeI* of modified pET14b.
pET3a-Cphy1180	PCR fragments of *Cphy1180* ligated within *NdeI* and *SpeI* of modified pET3a.
pET14b-Cphy1180	PCR fragments of *Cphy1180* ligated within *NdeI* and *SpeI* of modified pET14b.
pET3a-Cphy1181	PCR fragments of *Cphy1181* ligated within *NdeI* and *SpeI* of modified pET3a.
pET14b-Cphy1181	PCR fragments of *Cphy1181* ligated within *NdeI* and *SpeI* of modified pET14b.
pET3a-Cphy1182	PCR fragments of *Cphy1182* ligated within *NdeI* and *SpeI* of modified pET3a.
pET14b-Cphy1182	PCR fragments of *Cphy1182* ligated within *NdeI* and *SpeI* of modified pET14b.
pET3a-Cphy1184	PCR fragments of *Cphy1184* ligated within *NdeI* and *SpeI* of modified pET3a.
pET14b-Cphy1184	PCR fragments of *Cphy1184* ligated within *NdeI* and *BamHI* of modified pET14b.
pET3a-Cphy1186	PCR fragments of *Cphy1186* ligated within *NdeI* and *BamHI* of modified pET3a.
pET14b-Cphy1186	PCR fragments of *Cphy1186* ligated within *NdeI* and *SpeI* of modified pET14b.
pET3a-Cphy1176-1180-1181-1182-1184-1186	Engineered link-lock six shell proteins from *clostridium phytofermentans* ligated within *AatII* and *SpeI* of modified pET3a.
pET3a-Cphy1176-1180-1181-1182-1184	Engineered link-lock five shell proteins from *clostridium phytofermentans* ligated within *AatII* and *SpeI* of modified pET3a.
pET3a-Cphy1176-1180-1181-1182	Engineered link-lock four shell proteins from *clostridium phytofermentans* ligated within *AatII* and *SpeI* of modified pET3a.

pLysS-Cphy1176-1180-1181-1182	Engineered link-lock four shell proteins from *clostridium phytofermentans* ligated within *BsrGI* and *SpeI* of modified pLysS.
pET3a-Cphy1176-1180-1181	Engineered link-lock three shell proteins from *clostridium phytofermentans* ligated within *EcoRI* and *SpeI* of modified pET3a.
pET3a-Cphy1176-1180	Engineered link-lock two shell proteins from *clostridium phytofermentans* ligated within *EcoRI* and *SpeI* of modified pET3a.
pET3a-Cphy1184-1176	Engineered link-lock two shell proteins from *clostridium phytofermentans* ligated within *EcoRI* and *SpeI* of modified pET3a.
pET3a-Cphy1184-1182	Engineered link-lock two shell proteins from *clostridium phytofermentans* ligated within *EcoRI* and *SpeI* of modified pET3a.
pET14b-Cphy1186-1Ala	First Alanine mutation of *Cphy1186* optimized via QuikChange II XL Site-Directed Mutagenesis technique.
pET14b-Cphy1186-2Ala	Second Alanine mutation of *Cphy1186* optimized via QuikChange II XL Site-Directed Mutagenesis technique.
pGEM-T easy-Cphy1186	PCR fragments of *Cphy1186* ligated within 3´-T overhangs of modified Pgem-T easy.

2.1.4 Primers

All primers used in this study are listed in **Table 2.3**. Designed primers were ordered from Invitrogen Life Technologies, and Eurofins Genomics.

Table 2.3: List of designed primers used in this project.

Primer name	Sequence	Restriction enzyme site/ Amino acid codon	GC %	Length
Cphy1176-NdeI FW	5’CACCATATGACAACCGAAGATAAACTTCG 3’	NdeI	41	29
Cphy1176-SpeI RV	5’CACACTAGTTTAGGTACGTGTTATGTCACATAC 3’	SpeI	39	33
Cphy1180-NdeI FW	5’CACCATATGGGTTTAGCAGTTGGA 3’	NdeI	46	24
Cphy1180-SpeI RV	5’CACACTAGTTTAATTCTTATCAAAGCCGC 3’	SpeI	38	29
Cphy1181-NdeI FW	5’CACCATATGGGAAAATCGTTAGGC 3’	NdeI	46	24
Cphy1181-SpeI RV	5’CACACTAGTCTACATTTCTTCGTTTTGTGG 3’	SpeI	40	30
Cphy1182-NdeI FW	5’CACCATATGGATGATAAATTAGATAAAAAGTCG 3’	NdeI	30	33
Cphy1182-SpeI RV	5’CACACTAGTTTATTTAATTCCTGG 3’	SpeI	33	24
Cphy1184-NdeI FW	5’CACCATATGTTAATCGGCAAAGTAATCGGT 3’	NdeI	40	30

Cphy1184-SpeI RV	5'CACACTAGTTTACTCCAGTCCTGTTGC 3'	SpeI	48	27
Cphy1186-NdeI FW	5'CACCATATGAGTAAAGCAATTGGA 3'	NdeI	38	24
Cphy1186-SpeI RV	5'CACGGATCCCTACATCAACTTATTGCG 3'	BamHI	47	36
Cphy1186-1Ala FW	5' GCA CAA ACC GTT GCT CCG GGT AAA TAT ATT ATT TTA ATC 3'	Alanine	36	39
Cphy1186-1Ala RV	3' GAT TAA AAT AAT ATA TTT ACC CGG AGC AAC GGT TTG TGC 5'	Alanine	36	39
Cphy1186-2Ala FW	5' GCA CGT GGT ATG GCT GGT AAG TCT TAT CTT ATG 3'	Alanine	45	33
Cphy1186-2Ala RV	3' CAT AAG ATA AGA CTT ACC AGC CAT ACC ACG TGC 5'	Alanine	45	33

2.1.5 Media and solutions for culturing

Super Optimal Broth (SOB):

The media was sterilised by autoclaving. Glucose and $MgSO_4$ were filter sterilised. The pH was adjusted to 7 with 10 M NaOH.

Base medium	1000 mL
Tryptone (g)	20
Yeast extract (g)	5
NaCl (g)	0.5
1M KCl (mL, µL)	2.5
MQW (mL)	Up to 1000

Super Optimal Catabolite (SOC) repression broth:

20 mM glucose (3.603 g) were added to the 1 L SOB media before autoclaving. Both SOB or SOC media were used for bacterial transformations in order to provide the best growth for the competent cells.

Luria- Bertani (LB) broth:

This media was sterilised by autoclaving.

Base medium	1000 mL	200 mL
Tryptone (g)	10	2
Yeast extract (g)	5	1
NaCl (g)	5	1
1 M NaoH (mL, µL)	1	200
MQW (mL)	Up to 1000	Up to 200

Auto Induction Media (AIM) broth:

This media was sterilised by autoclaving. Glucose and $MgSO_4$ were filter sterilized.

Base medium	1000 mL	200 mL
Tryptone (g)	10	2
Yeast extract (g)	5	1
Na_2HPO_4 (g)	3.5	0.71
KH_2PO_4 (g)	3.4	0.68
NH_4Cl (g)	2.6	0.53
Na_2SO_4 (g)	0.7	0.14
Glucose (g)	0.5	0.1
Lactose (g)	2	0.5
$MgSO_4$ (g)	0.5	0.1
50 % Glycerol (mL)	10	2
MQW (mL)	Up to1000	Up to 200

Nutrient LB Agar:

15 g of bacterial agar were added to 1 L of nutrient LB agar and autoclaved (15 psi, 30 min).

Antibiotics:

Table 2.4: Antibiotics used for bacterial resistance.

Antibiotic name	Stock concentration	Working concentration
Ampicillin	100 mg/ mL	100 µg/ mL
Chloramphenicol	34 mg/ mL	34 µg/ mL

Isopropyl β-D-1-thiogalactopyranoside (IPTG):

9.53 g of IPTG were made up 20 mL to generate a 2 M IPTG solution, which was aliquoted into 1.5 mL Eppendorf tubes and stored at – 20°C. This solution was filter sterilised (0.2 µM).

$CaCl_2$ solution for competent cells:

100 mL of 100 mM $CaCl_2$ was used for the preparation of competent cells and minimal media. This solution was made up with distilled deionised H_2O and then autoclaved before used.

X-Gal stock solution 20 mg/ mL:

0.28 g X-Gal (5-Bromo-4Chloro-3-Indolyl-β-D Galactosidase) was dissolved in 14 mL DMF (Dimethyl formamide), wrapped with foil (light sensitive) and stored in – 20°C.

2.1.6 Media and solutions for DNA procedures

TE Buffer:	Tris-HCl, pH 8.0	10 mM
	EDTA pH 8.0	1 mM
6 x DNA Loading Buffer:	Glycerol	50 %
	TE Buffer	50 %
	Bromophenol blue	0.2 %

Molecular size marker HyperLadder™ 1kb:

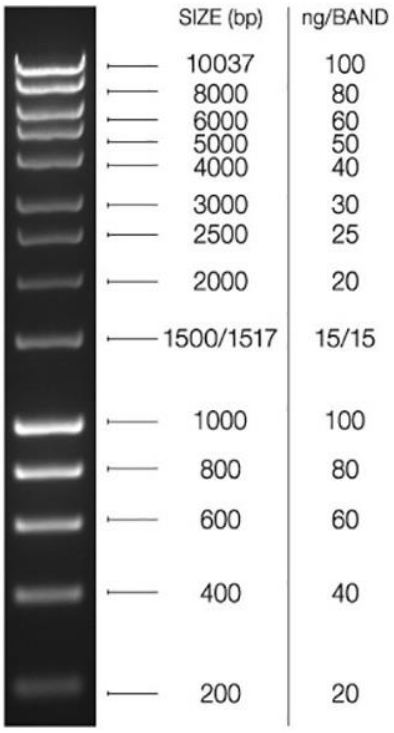


Figure 2.1: 5 µL HyperLadder™ 1kb run on a 1 % (w/v) Agarose gel in TAE.

2.1.7 Media and solutions for protein procedures

2.1.7.1 Solutions for protein purification

Charge buffer:	$NiSO_4$	100 mM
Binding buffer:	Tris-HCl, pH 8.0	20 mM
	NaCl	500 mM
	Imidazole	10 mM
Washing buffer I:	Tris-HCl, pH 8.0	20 mM
	NaCl	500 mM
	Imidazole	50 mM
Washing buffer II:	Tris-HCl, pH 8.0	20 mM
	NaCl	500 mM
	Imidazole	100 mM
Elution buffer:	Tris-HCl, pH 8.0	20 mM
	NaCl	500 mM

	Imidazole	400 mM
Strip buffer:	Tris-HCl, pH 8.0	20 mM
	NaCl	500 mM
	EDTA	100 mM
Desalting buffer (filtered):	Tris-HCl, pH 8.0	20 mM
	NaCl	100 mM
20 % Ethanol wash (filtered):	Ethanol	200 mL
	Distilled deionized H_2O	980 mL

2.1.7.2 Solutions for microcompartment purification

Solution A:	Reagent Y-PER™ (Yeast Protein Extraction Reagent is a mild detergent formulation that is superior to the classical methods of protein isolation from yeast)	10 mL
	Protease inhibitor (a tablet was dissolved in 1 mL distilled deionised H2O)	0.5 mL

	Benzonase Nuclease	5 µL
Buffer B:	Tris-HCl, pH 8.0	20 mM
	NaCl	20 mM

2.1.7.3 Solutions for SDS-PAGE

10 x Running buffer:	Tris-HCl	30 g/ L
	Glycine	144 g/ L
SDS gel 1 x Running buffer:	10 x Running Buffer	100 mL/ L
	10 % SDS	10 mL/ L
2 x Laemmli sample buffer:	0.5 M Tris-HCl, pH 6.8	2.5 mL
	10 % SDS	4 mL
	20 % Glycerol	2 mL
	10 % β-mercaptoethanol	1 mL
	0.08 % Bromophenol blue	500 µL
	Distilled H_2O	2.5 mL

14 µL β-mercaptoethanol added per mL 2 x SDS sample buffer.

Coomassie blue stain:	100 % Trichloracetic acid	250 mL
	Coomassie blue R250	0.6 g
	SDS	0.1 g
	Tris-HCl	0.25 g
	Glycine	0.15 g
	Distilled H_2O	Up to 500

Table 2.5: Gel composition for SDS-PAGE and 2D-PAGE.

SDS gels				
Running gels	**12.5 %**	**15 %**	**Stacking gel**	**15 %**
dH2O (mL)	3.4	2.2	**dH2O (mL)**	3.4
30 % Acrylamide (mL) Acryl/Bis™ 29:1	6.3	7.5	**30 % Acrylamide (mL) Acryl/Bis™ 29:1**	1.5
1.5 M Tris-HCl pH 8.8 (mL)	3.8	3.8	**0.5 M Tris-HCl pH 6.8 (mL)**	1.9
10 % (w/v) SDS (mL)	1.5	1.5	**10 % (w/v) SDS (mL)**	0.75
10 % (w/v) APS (mL)	0.15	0.15	**10 % (w/v) APS (mL)**	0.075
TEMED (mL)	0.01	0.01	**TEMED (mL)**	0.01

Pre-stained Blue Protein Standard, Broad Range:

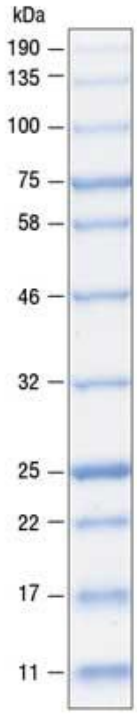


Figure 2.2: Pre-stained Blue Protein Standard, Broad Range is a mixture of highly purified, prestained proteins, covalently coupled with a blue chromophore that resolves into 11 sharp bands when electrophoresed. Ma, D. and Xu, MQ. New England Biolabs, Inc. Unpublished observation. Sambrook *et al*., 2001; Laemmli., 1970.

2.1.7.4 Solutions for EPR experiments

Binding buffer:	Tris-HCl, pH 8.0	20 mM
	NaCl	500 mM
	Imidazole	10 mM
Washing buffer I:	Tris-HCl, pH 8.0	20 mM
	NaCl	500 mM
	Imidazole	50 mM

Elution buffer:	Tris-HCl, pH 8.0	20 mM
	NaCl	500 mM
	Imidazole	400 mM
Desalting buffer (filtered):	Tris-HCl, pH 9.0	20 mM
	NaCl	100 mM
Sodium dithionite (sodium hydrosulfite):	$Na_2S_2O_4$	0.002 M

2.1.7.5 Solutions for Transmission electron microscopy experiments

Primary fixative:	2.5 % glutaraldehyde in 100 mM sodium cacodylate buffer, pH7.2 (cacodylic acid buffer, CAB).
CAB wash:	100 mM sodium cacodylate pH 7.2
Post fixative:	1% OsO4 (¼ (4 % OSO_4) + ¼ dH_2O + ½ (200 mM CAB))
Dehydration:	50 % Ethanol

	70 % Ethanol
	90 % Ethanol
	100 % Dried ethanol
	100 % Propylene oxide
Resin I:	½ Propylene oxide + ½ Resin II
Resin II:	24 g Viscosity Resin + 8 g VH1 + 18 g VH2 + 1.25 g LV accelerator
Staining stage I:	4.5 % Uranyl acetate in 1 % Acetic acid
Staining stage II:	0.1 % Reynold Lead citrate

2.2 Microbiological methods

2.2.1 Sterilisation

Prior to the start of any experiment, the working surface was sterilised with industrial methanol spirits. Tips, tubes, medias, and buffers were sterilised by

autoclaving for 15 min at 121°C and 1 bar pressure. Other liquids that are sensitive to high temperatures were filter sterilised (0.20 µm pore diameter). A flame was generally used to help maintain the sterility of bottled solutions.

2.2.2 Bacterial storage

All competent cells were incubated for 30 min on ice, then stored after the addition of glycerol to a final concentration of 15 % (v/v) at – 80°C. Pellets of bacteria containing expressed proteins were re-suspended with Binding buffer **(Section 2.1.7.1)** and then stored in – 80°C.

2.2.3 Solid cultures

Antibiotics and other supplements were added where required to solid media, which usually were spread from the transformation solutions, or streaked directly from the cells stock or source. Inoculated plates were incubated overnight at 37°C.

2.2.4 Liquid cultures

For plasmids and BMC isolation, a single colony was picked from a plate and was used to inoculate 5 mL LB. However, for protein expression many colonies were picked to inoculate a 10 mL LB starter culture, which, in turn, was used to inoculate 1 L baffled flasks. Cultures were incubated overnight at 37°C on a shaker at ~ 200 rpm. Antibiotics and other supplements were added as required.

2.2.5 Preparation of competent cells

Competent cells were prepared based on previously described methods (Sambrook *et al*., 1989). The required *E. coli* strain was streaked out into LB agar plates, which were then incubated overnight at 37°C. The following day, a single colony was used to inoculate a 5 mL LB starter culture that was incubated overnight at 37°C on a shaker at ~ 200 rpm. 1 mL of the starter culture was used to inoculate 100 mL SOB media and was grown to an OD_{600} of 0.4. The culture was cooled on ice for 15 mins and was then centrifuged at 3000 rpm for 10 min at 4°C. The pellet were resuspended in 50 mL ice-cold 0.1 M $CaCl_2$ and incubated on ice for 10 min prior to centrifugation. The cells were resuspended gently in 30 mL ice-cold 0.1 M $CaCl_2$ and incubated on ice for 30 min before being pelleted by centrifugation. Finally, the cells were gently resuspended in 4 mL of ice-cold 0.1 M $CaCl_2$ containing 15 % glycerol (v/v) and placed on ice for 10 min. The cells were next dispensed into 50 µL aliquots onto dry ice and were stored at – 80°C. Antibiotics and other supplements were added as required.

2.2.6 Transformation of competent cells

Competent cells were defrosted on ice for 10 min prior to the addition of 1 µL of plasmid or ligation mixture was added and incubated on ice for 15 min. The transformation mixture was heat-shocked for exactly 50 sec at 42°C, and was rapidly transferred to ice for 2 min. 200 µL SOC media with no antibiotic was added to the mixture and incubated for 30 - 40 min at 37°C

before it was spread on a LB agar plate containing the required antibiotics and incubated overnight at 37°C.

2.2.7 Production of recombinant protein

For protein production, in general, *E. coli* strain BL21 Star™ (DE3)pLysS was transformed with a pET plasmid containing a gene encoding a Hexa-His-tagged version of the protein of interest. The recombinant strains were grown in LB or AIM media with addition of Ampicillin and Chloramphenicol as required. Cultures were incubated at 37°C on a shaker until the OD_{600} reached 0.6. Protein expression was induced by the addition of 0.4 mM IPTG and the cells were left to grow overnight at 18°C on the shaker. Cultures then were centrifuged at 4000 rpm for 20 min at 4°C and pellets were re-suspended in Binding Buffer **(Section 2.1.7.1)**. The resuspended cells were transferred to a 50 mL Falcon tube and stored at – 80°C.

2.2.8 Lysis of cells using sonication

Harvested cells were lysed by sonication using a Sonics Vibracell Unltrasonic processor, with an output wattage of between 20 and 30 in 30 sec bursts with 30 sec breaks repeated six times. The sonicated cells were centrifuged at 35,000 x g for 20 min at 4°C to remove cell debris

2.3 Molecular biological methods

2.3.1 PCR reactions

The basic PCR procedures used in this study are outlined in **Table 2.6 & Table 2.7**. All PCR reactions were performed in an Eppendorf tube AG 22331. After reaction was completed, the PCR products were loaded onto an Agarose gel for separation by electrophoresis.

Table 2.6: Basic PCR reaction protocol and values.

	Reaction I	**Reaction II**
Distilled deionised H_2O	33.5 µL	38.5 µL
10 x PCR buffer (Roche, containing 18 mM $MgCl_2$)	5 µL	
5 mM dNTPs	1 µL	
DMSO	5 µL	-
10 µM 5' primer	2 µL	
10 µM 3' primer	2 µL	
DNA template/ 1 colony in 10 µL water	1 µL	
Roche Taq polymerase	0.5 µL	
Total	50 µL	

Table 2.7: Physical PCR reaction protocol.

Step	**Temp**	**Time**	**Cycles**	**Function**
		Roche *Taq*		
1	95°C	2 min	1	Initial Denaturation of the chromosomal DNA
2	95°C	30 sec	35	Denaturation of the amplified DNA
3	60°C	30 sec		Annealing of the primers
4	72°C	1 min/ 1000 bp		Elongation
5	72°C	5 min	1	Final elongation
Stop	4°C	Hold		

2.3.2 DNA Electrophoresis

DNA fragments were separated by agarose gel electrophoresis.

2.3.2.1 Agarose gel

Depending to the DNA fragments size, agarose gels of different percentages were chosen from 1 - 3 % (w/v) Agarose in 1 x TAE buffer. The agarose solution was dissolved by heating in a microwave oven and, after cooling, ethidium bromide was added to a final concentration of 0.5 µg/ mL. DNA loading buffer was added to DNA samples at 20 % (v/v) before they were loaded into the wells and running at ~ 110 V for an hour.

2.3.2.2 Visualisation under UV

The fluorescent dye ethidium bromide that was added to the agarose gel absorbs radiation (312 nm) of UV light and emits at 590 nm as visible orange bands when chelated to DNA. Thus, the gel was placed on the UV transilluminator to be photographed and saved.

2.3.3 DNA extraction

2.3.3.1 Extraction from Agarose gel

The gene of interest was carefully cut from the agarose gel with a clean scalpel blade and transferred into labelled Eppendorf tubes. The DNA was extracted from the gel according to the Qiagen QIAquik® gel extraction kit protocols.

2.3.3.2 Extraction from solutions

PCR products and digested DNA fragments were isolated using the Qiagen QIAquik® gel extraction kit by following the protocols described in the QIAquik® PCR purification handbook.

2.3.3.3 Extraction of plasmid DNA

To isolate plasmids, starter cultures were grown from single colonies and incubated overnight (antibiotics were added as required). Cultures were harvested by centrifugation at 4000 rpm for 5 min at 4°C. The cells were lysed and plasmids isolated by following the protocol described in the Qiagen QIAprep® miniprep kit.

2.3.4 Restriction enzyme digest

Restriction enzymes were frequently used to cut DNA at specific sites in order to generate specific DNA fragments for further cloning or to confirm the correct cloning of a gene. To perform a successful digestion, restriction enzymes and their optimal buffers were chosen carefully outlined in the information provided by the suppliers via either New England Biolabs or Promega. A standard protocol with slight changes was used to set the reactions **(Table 2.8)**. All the reactions were incubated at 37°C for two hours prior to electrophoresis.

Table 2.8: Typical restriction enzymes digestion protocol.

	Single digestion	**Double digestion**
Distilled deionised H_2O	3 µL	
10 x Buffer	1 µL	
Restriction enzyme 1	1 µL	0.5 µL
Restriction enzyme 2	-	0.5 µL
DNA	5 µL	
Total	10 µL	

2.3.5 DNA ligation

The plasmid and insert were both digested with relevant restriction enzymes in order to generate sticky compatible ends **(Section 2.3.4)**. A standard protocol with slight changes was used to initiate the ligation reactions **(Table 2.9)**. Reactions were incubated at room temperature for an hour or overnight at 4°C before being transformed into the appropriate competent cells **(Section 2.2.6)**.

Table 2.9: Typical DNA ligation protocol.

	The reaction
2 X Ligation Buffer	5 µL
T_4 DNA Ligase (Promega)	1 µL
Vector	1 µL
DNA insert	3 µL
Total	10 µL

2.3.6 Cloning into vectors

2.3.6.1 Cloning into pGEM®-T Easy-Indirect cloning

PCR products were ligated into the pGEM®-T Easy vector according to the *Quik* protocol provided by Promega **(Table 2.10)**. The ligation mixture was incubated for one hour at room temperature or overnight at 4°C before being transformed into *E. coli* competent JM109 cells **(Section 2.3.5)**. The cells were grown on LB agar plates including ampicillin, 0.4 mM IPTG, and 80 µg/mL X-Gal. Plates were incubated overnight at 37°C and the plasmid of

interest was identified by blue/ white colonies screening. The plasmids were isolated, digested with NdeI & BamHI and the gene was inserted within the compatible site in the relevant pET plasmid before being retransformed. To verify the DNA sequence, recombinant plasmids were test digested and sequenced.

Table 2.10: Promega *Quik* pGEM®-T Easy vector ligation protocol.

Standard reaction	
Distilled deionised H_2O	5 µL
10 x ligation buffer	1 µL
pGEM-T Easy vector	1 µL
DNA (Cphy_1186)	2 µL
T_4 Ligase	1 µL
Total	10 µL

2.3.6.2 Cloning into pET3a, pET14b and pLysS vectors

PCR products of genes of interest (*Cphy_1176, Cphy_1180, Cphy_1181, Cphy_1182,* and *Cphy_1184*) were digested with NdeI and SpeI, while *Cphy_1186* was digested with BamHI because of the naturally occurring SpeI site present within the gene **(Section 2.3.6.1)**. Digested genes were individually ligated within pET14b & pET3a plasmids that had also been digested with compatible enzymes. pET14b encodes an N-terminal Hexa-His tag, which is fused onto the encoded recombinant protein. The Cphy_1176 gene was ligated within the pLysS plasmid, after the plasmid been digested with compatible restriction enzymes. The plasmid was then used for the insertion of the *C. phytofermentans Cphy* genes Cphy_1180- Cphy_1181-

Cphy_1182, which were extracted from the plasmid pET3a-Cphy1176-1180-1181-1182. The resultant plasmids were transformed into competent cells. Single colonies were picked, and grown to amplification of the plasmid. To verify the DNA sequence, the recombinant plasmids were test digested and sequenced.

2.3.6.3 Multi-cloning strategy for the *C. phytofermentans Cphy* genes

All the *C. phytofermentans Cphy* six genes were ligated respectively in frame in pET3a by a technique called 'Link and Lock', a method developed by Dr. Evelyne Deery **(Figure 2.3)** (McGoldrick *et al*., 2005). Based on this strategy, plasmid pET3a was digested to allow for the cloning of the first *C. phytofermentans* gene of interest **(Section 2.1.4)** using the SpeI and EcoRI sites. The second *C. phytofermentans* gene of interest, was digested with the compatible cohesive restriction sites EcoRI & XbaI. The ligation of the SpeI & XbaI results in the loss of both sites and thereby produces a new plasmid with the required genes with just one SpeI site.

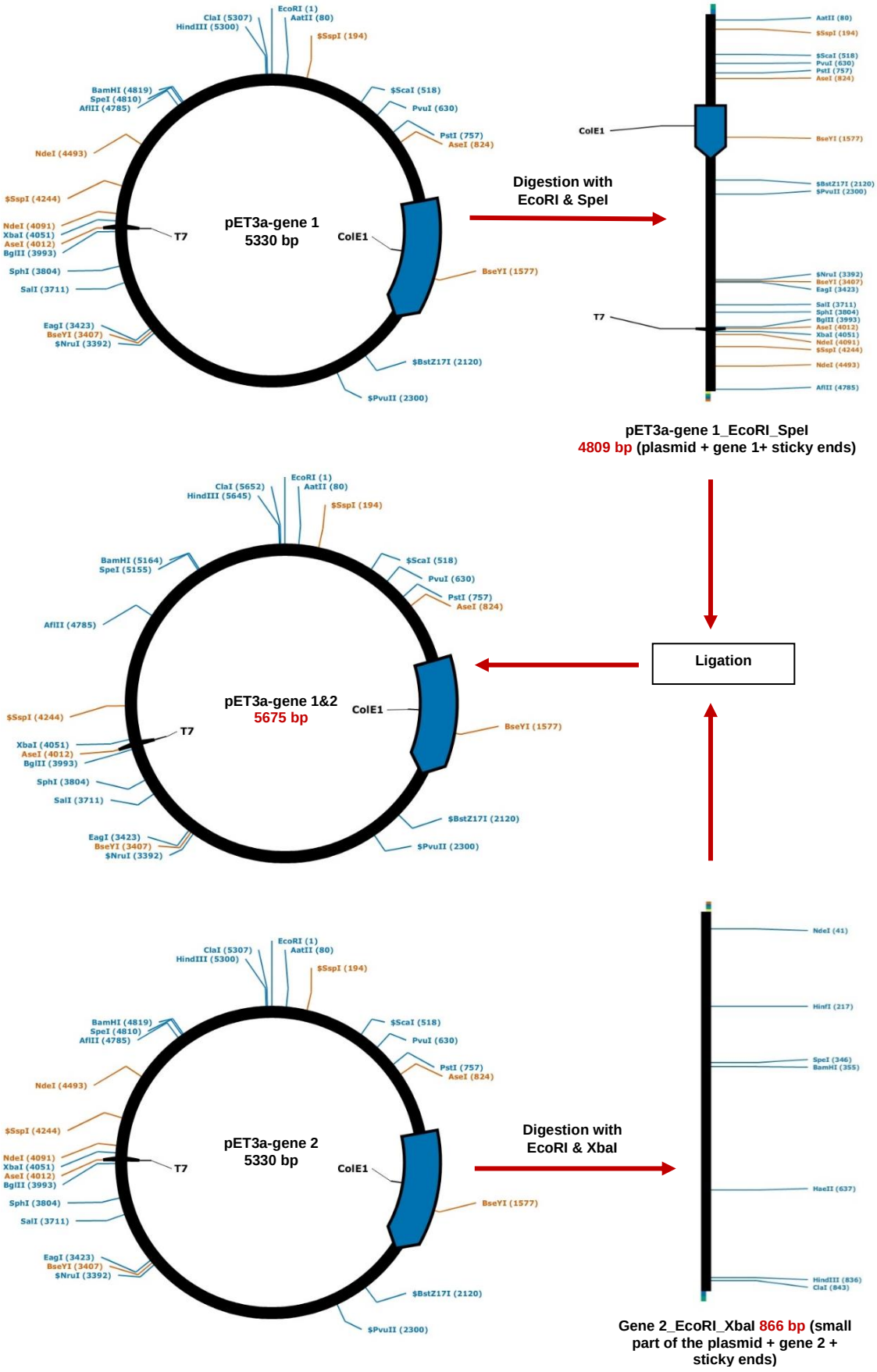


Figure 2.3: Schematic drawing of 'Link and Lock' cloning.

2.3.6.4 Mutagenesis by overlap extension of *Cphy_1186*

Two individual mutations in *C. phytofermentans* Cphy1186 (C_{38} and C_{133}) were generated using the QuikChange® II site-directed mutagenesis kit, a method provided by Stratagene. **Figure 2.4** summarises this technique that creates point mutations that change, delete or inserts single or multiple base pairs. The main procedure utilises a supercoiled double-stranded DNA (dsDNA) plasmid with the insert of interest and two synthetic oligonucleotide primers, both of which include the desired mutation. The oligonucleotide

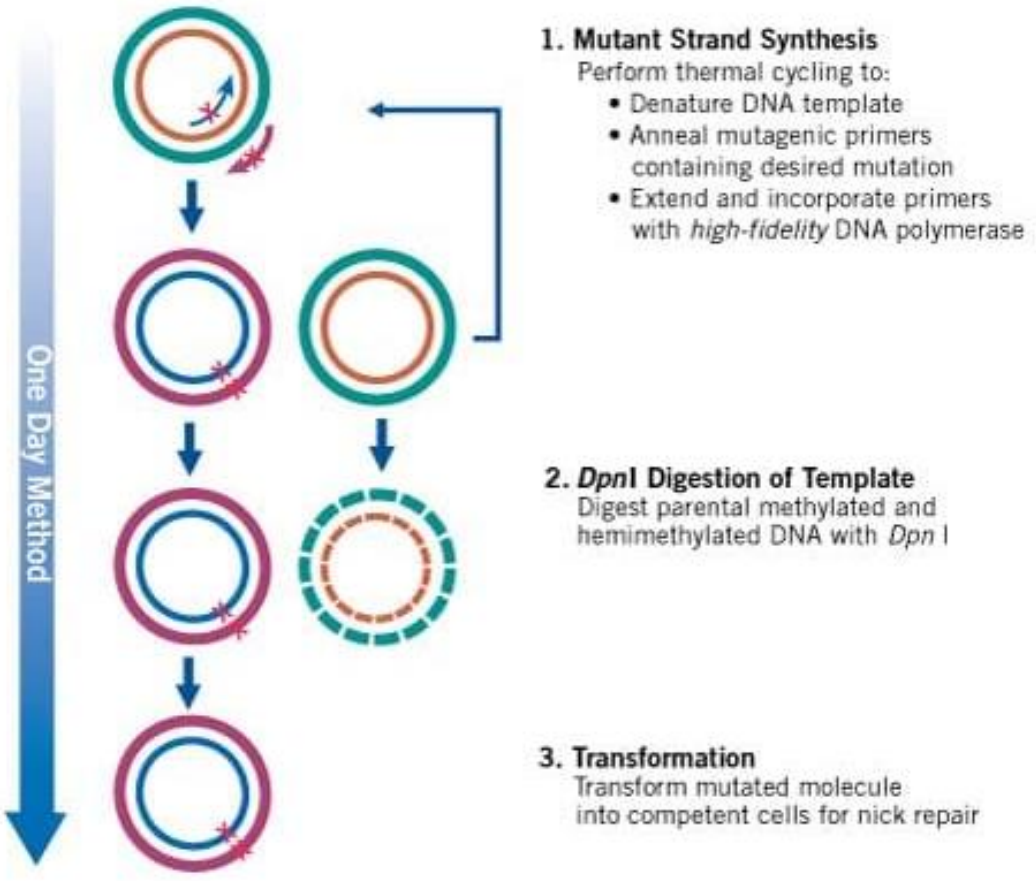


Figure 2.4: Overview of the QuikChange II site-directed mutagenesis method. This figure derived from the Instruction Manual of QuikChange® II site-directed mutagenesis kit.

primers, each complementary to opposite strands of the plasmid, were extended during temperature cycling via PfuUltra^{TM} HF DNA polymerase. The oligonucleotide primers extensions create a mutated plasmid containing staggered nicks. Following temperature cycling, the products were treated with DpnI. This endonuclease enzyme (target sequence: 5´-Gm6ATC-3´) is specific for methylated and hemimethylated DNA, to digest the parental DNA template and to select for mutation-containing synthesized DNA. The nicked plasmid DNA including the desired mutations was then transformed into XL1-Blue super competent cells. Finally, the plasmids were isolated and sequenced.

2.4 Biochemical methods

2.4.1 Standard purification procedures

2.4.1.1 Immobilised metal affinity chromatography

An 8.3 mL size syringe was filled with 4 mL Chelating Sepharose^{TM} Fast Flow resin that was charged with $NiSO_4$ (Charging buffer). The column was washed with milliqu water and equilibrated with 50 mL Binding buffer before the His-tagged protein from 1 L culture being loaded. As this kind of fused protein has affinity to the transition metals such as Ni^{2+}, Fe^{2+}, and Cu^{2+}, therefore it will retain with exposed histidine, which form a complex with the metal ions. At room temperature, the bound protein then was eluted by a high concentration of imidazole buffer. For cell pellet lysis, **Figure 2.5** summarise it. Supernatant, where the protein located, was loaded into the charged column and allowing to flow through via gravity, as well as washed

with 50 mL Binding buffer, 25 mL Washing buffer I, and 25 mL Washing buffer II. The bound protein then was isolated via 15 mL Elution buffer and the eluted fractions were collected in 2 mL one by one **(Figure 2.6)**. Before storing the column at 4°C, resin was regenerated with 15 mL Strip buffer then washed with 50 mL water. The composition of purification buffer can be found in **Section 2.1.7.1**.

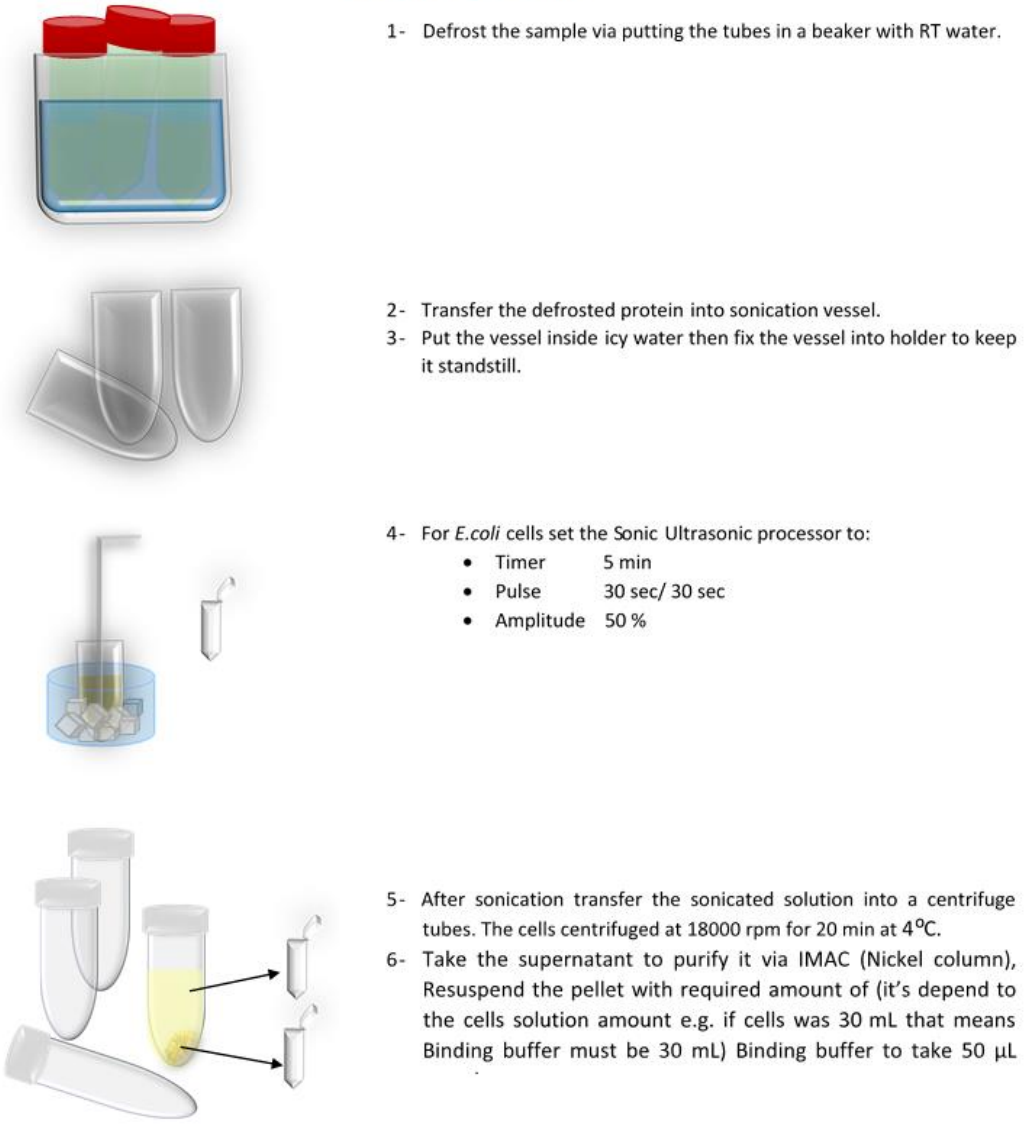


Figure 2.5: Schematic drawing of cell lysis technique.

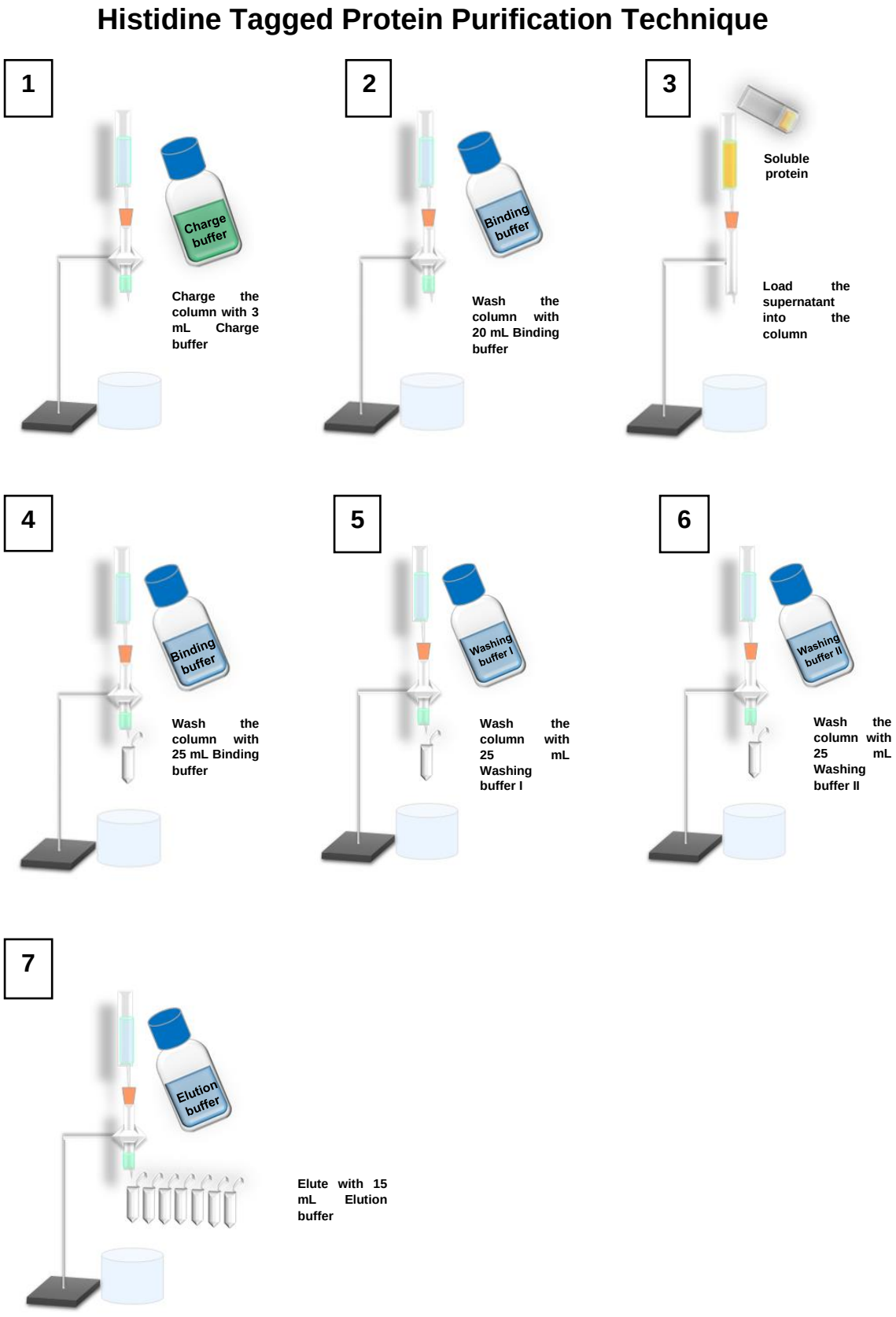


Figure 2.6: Schematic drawing of protein purification technique (IMAC).

2.4.2 Buffer exchange

Purified protein was desalted using pre-packed disposable Sephadex G25 columns (PD-1, bed volume 8.3 mL). The column tip was cut off, liquid removed, and washed with 50 mL distilled deionised H_2O. The resin was equilibrated with 25 mL desalting buffer. Protein was applied to the column in a volume of exactly 2.5 mL, and allowed to flow through via gravity. Finally, the protein was eluted in a volume of 3.5 mL elution buffer. The collected fractions were analysed by SDS-PAGE. The composition of purification buffer can be found in **Section 2.1.7.1**.

2.4.3 Calculation of protein concentration

The concentration of purified desalted protein was estimated via the Warburg and Christian method (Warburg, 1941). The OD_{280} was recorded and the extension coefficient estimated by the computational ProtParam tool on the ExPASy® server. Protein concentration (Molar) = $\frac{A_{280}}{\text{Extension coefficient}}$

mg/ mL protein concentration = Molecular mass x Concentration (Molar)

2.4.4 Fast protein liquid chromatography

Gel filtration was achieved on an Amersham Biosciences P920 Fast Protein Liquid Chromatography (FPLC) system using a Superdex^{TM} G200 column (GE Healthcare). The column was washed with distilled deionised H_2O and pre-equilibrated with 1.2 column volumes of desalting buffer **(Section 2.1.7.1)**. The purified protein was concentrated to a volume of 1 mL and

injected into the column. Protein fractions were collected at a flow rate of 0.5 mL/ min and analysed by SDS-PAGE. Finally, the column was washed with 1 column volume of water containing 20 % ethanol.

2.4.5 Polyacrylamide gel electrophoresis

2.4.5.1 SDS-PAGE

Protein samples of interest (50 µM) were collected and denatured by the 1:1 addition of 2 X Laemmli buffer, and boiling for 5 min. The recipe of the SDS gel, buffers, and stain can be found in **Section 2.1.7.3**. Protein samples (5 - 15 µL) were loaded into individual wells of the gel together with 5 µL of the molecular mass marker Dalton VII. The protein standards are run on the gel in order to know the relative molecular mass of the protein of interest. Protein samples were analysed using the sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) based on a method developed by Laemmli (Laemmli, 1970). Proteins samples were run for 90 min at 200 V using a BioRad Mini Sub® Cell GT electrophoresis apparatus, powered from a BioRad Power PAC 300 power supply. After electrophoresis the gels were stained for 15 – 30 min with coomassie blue stain, and de-stained with distilled water.

2.4.6 Anaerobic techniques

Some procedures, such as Fe-S protein preparations, were carried out anaerobically under a nitrogen environment, which contains less than 3 ppm oxygen. In this case, all the chemicals were brought into the glove box (Belle

Technologies) as powders, and liquids were degassed by sparging with argon gas before use. Nickel columns were charged and washed out of the glove box. The columns were then transferred into the glove box 24 hours prior to use and equilibrated with 50 mL of the required buffers.

2.4.7 Microcompartment purification

The recombinantly expressed BMCs were extracted from cells using yeast protein extraction reagent (Y-PER™). *E. coli* BL21 Star™ (DE3)pLysS cells were transformed with plasmid pET3a-Cphy_1176, 1180, 1181, 1182, 1184, 1186 and grown in 500 mL baffled flasks containing 250 mL of LB media with addition of appropriate antibiotics, ampicillin and chloramphenicol. The starter culture was a 5 mL LB culture grown at 37°C with shaking at 250 rpm. One of these transformants was grown and incubated until an OD_{600} reading of approximately 0.6 was achieved. At this point protein overproduction was induced by the addition of IPTG to induce expression of T7 RNA polymerase and grown at 18°C overnight. After overnight growth the culture were harvested by centrifugation at 2700 x g (3000 rpm) at 4°C for 10 minutes and any macromolecular protein complex that had formed was purified using Y-PER™. The culture was subject to the purification protocol that had been developed for recombinant empty BMCs from *E. coli*. The cell pellet was resuspended in solution A **(Section 2.1.7.2)**, and the suspension was incubated at room temperature with shaking at 60 rpm for 3 hours. Subsequently, debris was removed by centrifugation at 12,000 x g at 4°C for 5 minutes. The pellet was resuspended in 3 mL of buffer B **(Section 2.1.7.2)**, and centrifuged at 12,000 x g at 4°C for 5 minutes. The resulting supernatant

was centrifuged at 12,000 x g at 4°C for 5 minutes after the addition of 5 M NaCl in order to raise the salt concentration to 80 mM. Finally, the pellet was resuspended in 200 µL of 20 mM Tris-HCl, pH 8.0, and the resulting suspension containing purified BMCs was stored on ice prior for analysis.

2.4.8 Preparation of samples for transmission electron microscopy

First day: *E. coli* BL21 Star™ (DE3)pLysS cells were transformed with the plasmid pET3a containing the required gene.

Second day: One colony was picked to inoculate 5 mL LB media with the addition of the appropriate antibiotics, ampicillin as well as chloramphenicol and grown at 37°C with shaking at 250 rpm overnight.

Third day: 50 mL LB media with the addition of appropriate antibiotics, ampicillin as well as chloramphenicol was inoculated with 500 µL of the starter culture. The culture was incubated at 37°C with shaking at 250 rpm until the OD_{600} reading was approximately 0.6. At this point protein overproduction was induced by the addition of 400 µM IPTG and incubated for a further 2 hours. The cells were harvested by centrifugation at 3000 rpm at 4°C for 10 minutes in preparation for fixation. The cell pellet was resuspended in 2 mL fix consisting of 2.5 % glutaraldehyde in CAB and rotated at 4°C overnight **(Section 2.1.7.5)**.

Fourth day: The cells were centrifuged for 2 minutes at 10000 rpm and the pellets were washed twice with CAB to remove traces of the fixing solution.

The cells were stained for 1 hour in 1 mL of 1 % osmium tetroxide and then washed twice with MilliQ water before dehydration **(Section 2.1.7.5)**. Dehydration was accomplished by subjecting the cells to a solvent gradient: 50 % ethanol, 70 % ethanol rotated at 4°C overnight.

Fifth day: The dehydration was completed with 90 % ethanol, and 100 % dried ethanol for 10 minutes as well as twice further with 100 % dried ethanol for 2 hours. The cells were twice rinsed with propylene oxide for 10 minutes. The cells were pelleted and resuspended in 1.5 mL of (½ propylene oxide + ½ Agar Low Viscosity Resin) and rotated for 30 minutes **(Section 2.1.7.5)**. The cells were pelleted and resuspended twice in 1.5 mL of 100 % Agar Low Viscosity Resin for 1.5 hours. The cells were pelleted and resuspended in 0.5 – 1 mL 100 % Agar Low Viscosity Resin. The suspension was placed in 0.5 mL embedding tubes and centrifuged at 1100 rpm for 5 minutes. The tubes were incubated at 60°C for around 12 hours to polymerise.

2.4.9 Sectioning and staining of samples

Specimens were thin sectioned by Ian Brown using a diamond knife on an RMC MT-6000-XL ultramicrotome. The thin sections were collected on copper grids, which were stained with 4.5 % uranyl acetate in 1 % acetic acid for 45 minutes at room temperature. The samples were next stained with Reynold lead citrate for 7 minutes. Sections were observed and photographed with a JEOL-1230 transmission electron microscope.

2.4.10 Electron paramagnetic resonance spectroscopy (EPR)

With Dr. Steve Rigby (University of Manchester), EPR spectra was recorded. Samples were prepared anaerobically as previously described and stored in liquid nitrogen. EPR was performed on a Bruker ELEXSYS E500 spectrometer operating at X-band and employing a Super High Q cylindrical cavity (Q factor~16,000) equipped with an Oxford Instruments ESR900 liquid helium cryostat associated to an ITC503 temperature director. The iron-sulphur centres of *C. phytofermentans* Cphy1186, C38A and C133A were examined at 12 K employing a microwave power of 1mW, modulation frequency of 100 kHz, as well as modulation amplitude of 5 G.

2.4.11 Protein crystallisation

Crystallography is an experimental science that is used to determine the atomic arrangement in crystals in order to establish structure. Normally, to form good crystals proteins need to be as pure as possible and highly concentrated as well as homogenous. Once the protein becomes supersaturated crystals will emerge and grow. To reach the required crystallising condition, the solubility of the purified protein solution in a given set of conditions is gradually lowered through using a precipitant. There are a large number of factors affecting crystallisation which include; protein concentration, buffer, pH, precipitant, temperature and drop volume. The rate at which a protein solution becomes supersaturated is critical as if this happens too quickly often the protein would aggregate and form amorphous precipitates. Under the appropriate conditions, which is different for each protein, individual protein molecules align themselves in a repeating series of unit cells and thus forming crystals. **Figure 2.7** displays the relationship

between the concentration of the protein and the precipitant and the likely zone for crystal growth .

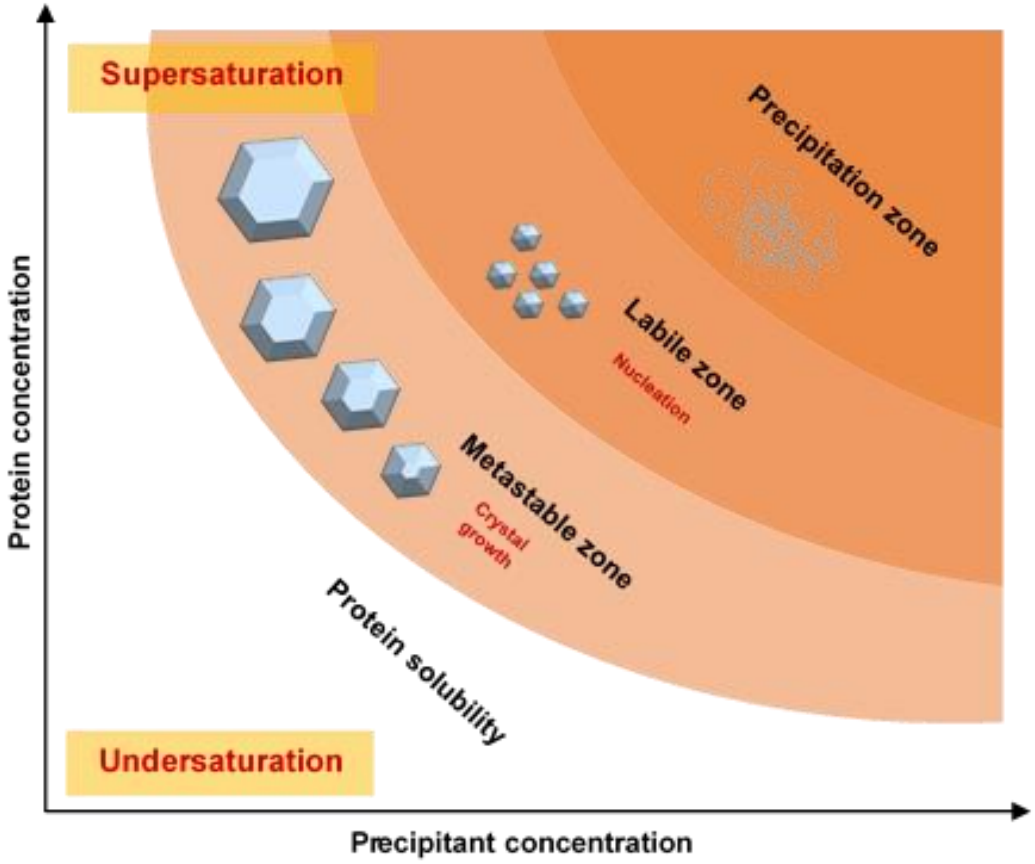


Figure 2.7: Schematic drawing of crystallisation phase diagram. As precipitant concentration increases, soluble protein enters the labile zone where nucleation occurs. Once nucleation happens, the concentration of protein lowers and the system enters the metastable zone, on which protein crystal forms.

For initial screening, protein samples were crystallised using the Molecular Dimensions Structure Screen I MD1-01 and II MD1-02 by use of the hanging drop method **(Figure 2.8)**. 500 µL crystal screen reservoirs were dispensed onto the 24 well-plate. 1 µL of the purified protein was mixed with 1 µL of the

crystal screen reservoir in the 22 x 22 mm coverslip (Menzel Glaser) to make one hanging drop. The coverslip was placed on top of the well and each well was sealed with a Dow Corning high vaccum grease. The trays were incubated at 18 ºC and checked regularly for crystal growth. Conditions that formed crystals in the initial screening, were optimised in attempt to get better crystals for X-ray. The optimisation procedure involves slight alteration of the reservoir (buffer pH and concentration, precipitant concentration and salt concentration).

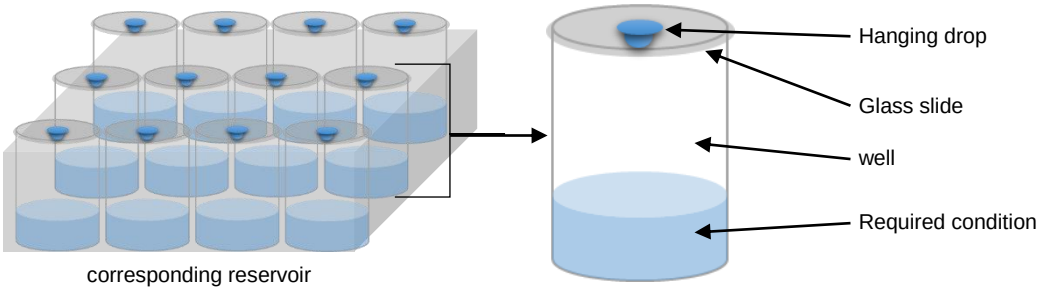
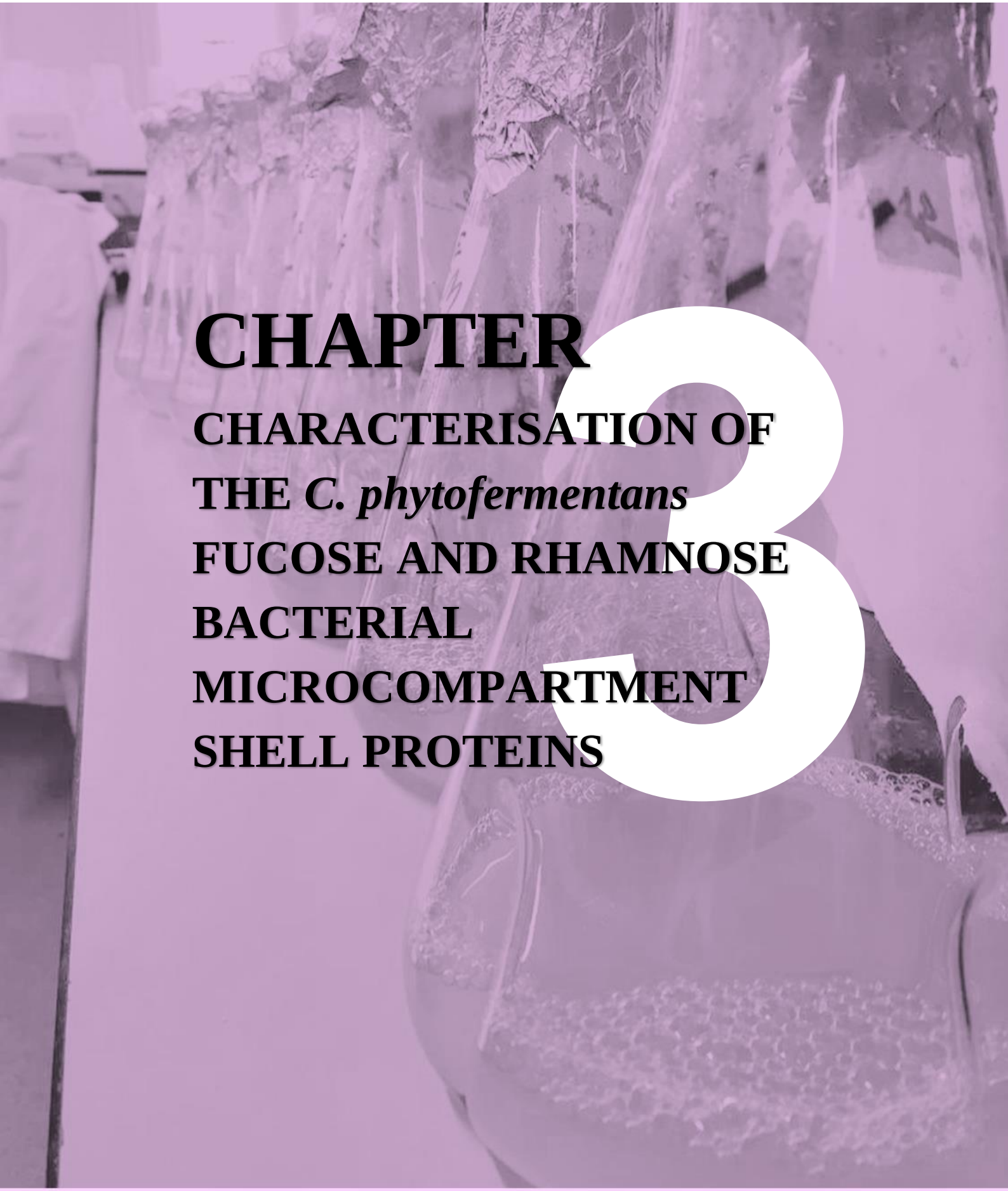


Figure 2.8: Schematic drawing of the hanging drop method. 1 µL of concentrated protein solutions mixed with 1 µL of reservoir solution in a drop on a coverslip, inverted over the reservoir solution.



CHAPTER 3

CHARACTERISATION OF THE *C. phytofermentans* FUCOSE AND RHAMNOSE BACTERIAL MICROCOMPARTMENT SHELL PROTEINS

3.1 Introduction

This chapter describes the initial steps in the characterisation of the *Clostridium phytofermentans* (Cphy) bacterial microcompartment. The wild type *C. phytofermentans* BMC operon has been previously reported and has been implicated in the growth of the organism on either fucose or rhamnose (Petit *et al.*, 2013). The operon is shown diagrammatically in **Figure 3.1**, and includes six shell protein genes (Cphy_1176, 1180, 1181, 1182, 1184, 1186), which are thought to form the outer structure of the polyhedral organelle. These six shell proteins, which were identified on the basis of sequence similarity to other known shell proteins also contain the Pfam motifs PF00936 and PF03319 (Petit *et al.*, 2013). Building on this genome analysis, and in order to help in the characterisation of the potential shell proteins associated with this bacterial microcompartment, the first aim of this project was to clone and express the individual six shell proteins of *C. phytofermentans* in *E. coli*, in order to purify and characterise the recombinant proteins produced. The second aim was to engineer an empty BMC through the coordinated production of the six shell proteins and to characterise the recombinant structure that was formed. In so doing, this research would generate more

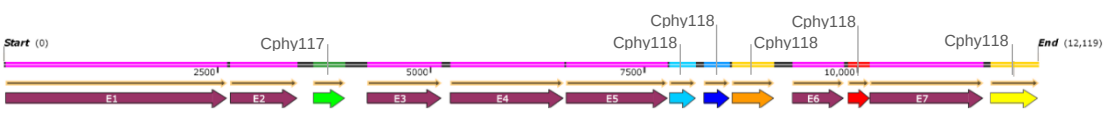


Figure 3.1: *C. phytofermentans* ISDg (Accession number NC_010001) fucose/ rhamnose utilisation 13-genes operon without the acetate kinase Cphy1327 and Cphy1187 the transcriptional regulator from DeoR family. Shell proteins are coloured and named, while the rest of metabolic process genes are in purple.

information about BMC composition and the minimum number of shell proteins necessary for empty microcompartment assembly.

3.2 Results

3.2.1 Sequence analysis

The *C. phytofermentans* ISDg genome, which can be found at http://www.ncbi.nlm.nih.gov, and its encoded protein sequences were analysed using blast searches (http://blast.ncbi.nlm.nih.gov/Blast.cg). Information about *C. phytofermentans* and its genome is presented in **Table 3.1**.

Table 3.1: Some related information of *C. phytofermentans* ISDg genome.

Molecular type	Nucleic acid
Query length	12119 bp
Accession	NC 010001.1
GC %	39 %
Kingdom	Bacteria
Phylum	Firmicutes
class	Clostridia
order	Clostridiales
Family	Clostridiaceae
Genus	Clostridium
Species	phytofermentans
Strain	ISDg
Sequencing centre	(08-NOV-2007) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA/ (21-NOV-2007) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA
Isolation site	Forest soil near the Quabbin Reservoir in Massachusetts
Isolation country	USA
Gram staining properties	Positive
Metabolism	Acetate producer/ Cellulose degrader/ Ethanol production

The DNA and amino acid sequence for each of the six shell proteins is displayed in **Table 3.2**, whereas information about each of the encoded proteins found within the operon is shown in **Table 3.3**.

Table 3.2: DNA and amino acids sequences of each shell protein in *C. phytofermentans*.

Protein's name	DNA sequence	Amino acids sequence
Cphy_1176	ATGACAACCGAAGATAAACTTCGCATTGTACAGGAATTAGTTC CTGGTAAGCAGATATCTCTGGCTCACATAATAGCTGCACCAG ATCCACTTCTAATGGAGAAATTGTGTTTTCAAAAAGAAAACGA GAGGATGAAGGCTGCTATCGGAATATTAACCATGAGCCCTGC GGAGACAGTAATTATTGGAGCTGATTTAGCATTGAAGGCATC GGGTGTTACACTGCAAAATGTAGATTATGCCAGCGGCACTTT AGTATTTACAGGAACAGTATCTGAGGTAGAAGCTGCGATGAA TGCGGTTGTGGAGTATTCAAATCGCACCTTGTCTTTTACTGTA TGTGACATAACACGTACCTAA	M T T E D K L R I V Q E L V P G K Q I S L A H I I A A P D P L L M E K L C F Q K E N E R M K A A I G I L T M S P A E T V I I G A D L A L K A S G V T L Q N V D Y A S G T L V F T G T V S E V E A A M N A V V E Y S N R T L S F T V C D I T R T
Cphy_1180	ATGGGTTTAGCAGTTGGATTTTTAGAGGTATATGGTTTGACGG CAGCATTTGTGGCAGCCGACGCCGGATGTAAGGCCGCTGAT GTAACCTTGGAACCATTTGACCGAAATAAGCCGGCAAATGCA GATAAACTTCCCGTTCCATTATTGGTTACGATAAAATTCAGAG GTTCCATTGAGGATGTCAATGCAGCTTTAGATGCAGCGAAATT AGCTGCAAAGCAAGTTACCGATGTCGTATCCATCCATACAATA GCCGCACCTGAGGGTGATGTGAATAAAATGCTTCACTTAAGC GGCTTTGATAAGAATTAA	M G L A V G F L E V Y G L T A A F V A A D A G C K A A D V T L E P F D R N K P A N A D K L P V P L L V T I K F R G S I E D V N A A L D A A K L A A K Q V T D V V S I H T I A A P E G D V N K M L H L S G F D K N
Cphy_1181	ATGGGAAAATCGTTAGGCTTTATTGAAATCAGTGGTGTAACCG CAGCGATTGATGCCCTCGATATTATGTGTAAAACAGCAGGTG TCGAGTTAGTTACATGGGAGAGAAAGATGGGTGGAAGGTTGG TTACAATCATCGTATCAGGATCTGTTTCCGAAGTAACACAGGC AGTAGAATCAGCGGTAAATCAGGCAATTAAAAAACCAGTGGC CCATGCAGTCATTGCAAATCCTCATGAAGAGGTTCTTCGCTTG GTAGATATCAGTGCAAGTAGAATGAAGCCACAAAACGAAGAA ATGTAG	M G K S L G F I E I S G V T A A I D A L D I M C K T A G V E L V T W E R K M G G R L V T I I V S G S V S E V T Q A V E S A V N Q A I K K P V A H A V I A N P H E E V L R L V D I S A S R M K P Q N E E M
Cphy_1182	ATGGATGATAAATTAGATAAAAAGTCGAATTCGACGAGAAAAG CTGAGATTACTGGAGAGGGAAAAAAGTCTCGCAGTACATCCA AAGTATCACTTACCGGACAAAGCATTGATGCGGTAAGCGGTA GCAACATTTCCATGGAAGTTATGAAAGCTAATAAAAATATAAA TGAAGAGGTTACTGTTGATACAACAGTTCTTAAGGAGGAAAAA ATTATGGCACAGGAAGCATTAGGAATGATCGAAACAAGAGGT TTAGTAGCAGCGGTAGAGGCAGCAGATGCAATGTTAAAGGCA GCTAACGTAACATTAGTAGGAACAGAAAAAATCGGTTCTGGAT TAGTAAGTGTTATGGTTAGAGGAGATGTAGGAGCAGTGAAAG CAGCTGTTGAGGCTGGTGCAGCTAGCGCTGGAAGACTTGGT GAGTTAGTAGCTACTCATGTAATCCCAAGACCACACTCTGATG TTGAGAAAATTTTACCAGGAATTAAATAA	M D D K L D K K S N S T R K A E I T G E G K K S R S T S K V S L T G Q S I D A V S G S N I S M E V M K A N K N I N E E V T V D T T V L K E E K I M A Q E A L G M I E T R G L V A A V E A A D A M L K A A N V T L V G T E K I G S G L V S V M V R G D V G A V K A A V E A G A A S A G R L G E L V A T H V I P R P H S D V E K I L P G I K
Cphy_1184	ATGTTAATCGGCAAAGTAATCGGTAGTGTTGTTTCAACAAGAA AGAATGAGAATCTGGTTGGAAATAAATTTATGATTGTTGAACC TTTAAAAAGCTTTCAAGAAAATAGAATAGTAGCAATTGATAATG TAGGAGCCGGAATAGGTGAATATGTGTTAGTTGCACAGGGTA GTGCTGCAAGAATAGGCTGCGGTGCAGAAAACTCGCCGATT GATTCGGCAATTGTTGGAATTATTGACGATGCAACAGGACTG GAGTAA	M L I G K V I G S V V S T R K N E N L V G N K F M I V E P L K S F Q E N R I V A I D N V G A G I G E Y V L V A Q G S A A R I G C G A E N S P I D S A I V G I I D D A T G L E
Cphy_1186	ATGAGTAAAGCAATTGGAATGGTAGAATATAAAACCGTATCAT CAGGTATTATGGCAGCGGATTTGATGGTTAAAACAGCAGATA TTGACATCGTGGAAGCACAAACCGTTTGTCCGGGTAAATATAT TATTTTAATCACTGGTGATTTAAGTGCGGTAAATGCTTCGGTA GAGGCAGCCAGAATTCAGTTTGAAACACATCTCATTGATAGTT TCATACTTGGTAATCCTCACGATGGTATTTTACCGGCAATTTA CGGTGCGTCAGTGGTGGAAGAAATCGAAGCGCTTGGTGTGT TAGAAACTTACTCAGCAGCTTCCATTATCGTTGCGGCTGATGT TGCAGCAAAAACAGCAGCTGTTGAACTAGTAGAGGTTAGGGT CGCACGTGGTATGTGTGGTAAGTCTTATCTTATGTTAACCGGT GAAATTGCATCAGTAACCGCTTCTATTGAAGCTGCAAAGAAAG CCATTGGAGAAAACGGTATGTACTTAGATAGCTCTGTACTTGC ACATCCAGATAAGAAGTTACGCAATAAGTTGATGTAG	M S K A I G M V E Y K T V S S G I M A A D L M V K T A D I D I V E A Q T V C P G K Y I I L I T G D L S A V N A S V E A A R I Q F E T H L I D S F I L G N P H D G I L P A I Y G A S V V E E I E A L G V L E T Y S A A S I I V A A D V A A K T A A V E L V E V R V A R G M C G K S Y L M L T G E I A S V T A S I E A A K K A I G E N G M Y L D S S V L A H P D K K L R N K L M

Table 3.3: genetic organisation and some functional equivalent of *C. phytofermentans* fucose/ rhamnose bacterial microcompartment.

***C. phytofermentans* genes name**	**Start**	**End**	**Length (bp)**	**Functional equivalent *S. enterica***	**Size (kDa)**	**% Total spectra**	**Proposed function**
Cphy_1174	1,489,322	1,491,925	2,604	PduCDE	97	0.087	Propanediol dehydratase
Cphy_1175	1,491,961	1,492,752	792	PduGH	29.62	-	Propanediol dehydratase activator
Cphy_1176	1,492,951	1,493,310	360	PduU	12.9	0.0066	Shell protein (PF00936)
Cphy_1177	1,493,642	1,494,436	795	nd	29	0.035	Fucolose and rhamnulose phosphate aldolase
Cphy_1178	1,494,484	1,495,872	462	PduP	49.57	-	Propionaldehyde dehydrogenase
Cphy_1179	1,495,895	1,497,094	1200	PduQ	43	0.11	Propanol dehydrogenase
Cphy_1180	1,497,110	1,497,424	315	$PduA_1$	10.9	0.0058	Shell protein (PF00936)
Cphy_1181	1,497,514	1,497,816	303	PduK	10.7	0.0041	Shell protein (PF00936)
Cphy_1182	1,497,843	1,498,337	495	$PduA_2$	17	0.043	Shell protein (PF00936)
Cphy_1183	1,498,518	1,499,159	642	PduL	23.06	-	Phosphate propanoyl transferase
Cphy_1184	1,499,200	1,499,460	261	PduN	9	0.017	EutN_CcmL shell protein (PF03319)
Cphy_1185	1,499,464	1,500,792	1329	PduS	48	0.013	Propanediol oxidoreductase
Cphy_1186	1,500,877	1,501,425	549	PduT	19.2	0.0058	Shell protein (PF00936)
Cphy_1327	1,658,810	1,659,997	1188	-	29	0.0025	Acetate/propionate kinase

3.2.2 Amplification and cloning of genes encoding the six shell proteins

Based on the *C. phytofermentans* 13-gene BMC operon shown in **Figure 3.1** and **Table 3.3**, the six putative shell-protein genes (termed Cphy_1176, 1180, 1181, 1182, 1184, 1186) were amplified by PCR **(Figure 3.2)** using the primer sequences shown in **Section 2.1.4**. The primer sequences were designed to contain appropriate restriction sites to allow cloning of the genes individually into pET14b and collectively into pET3a. The PCR products of the amplification were digested with the appropriate restriction enzymes in order to generate sticky ends ready for the subsequent ligation into the pET plasmids. As mentioned previously in **Section 2.1.3**, all amplified and digested genes were ligated into the plasmids pET3a and pET14b **(Figure 3.3 and 3.4)**. However, because of an endogenous *SpeI* site within one of the shell protein genes, Cphy_1186 was cloned directly into pGEM-T vector and then subcloned into pET3a and pET14b vectors **(Figure 3.5)**. All the inserts were sequenced by Genewiz, UK before expression.

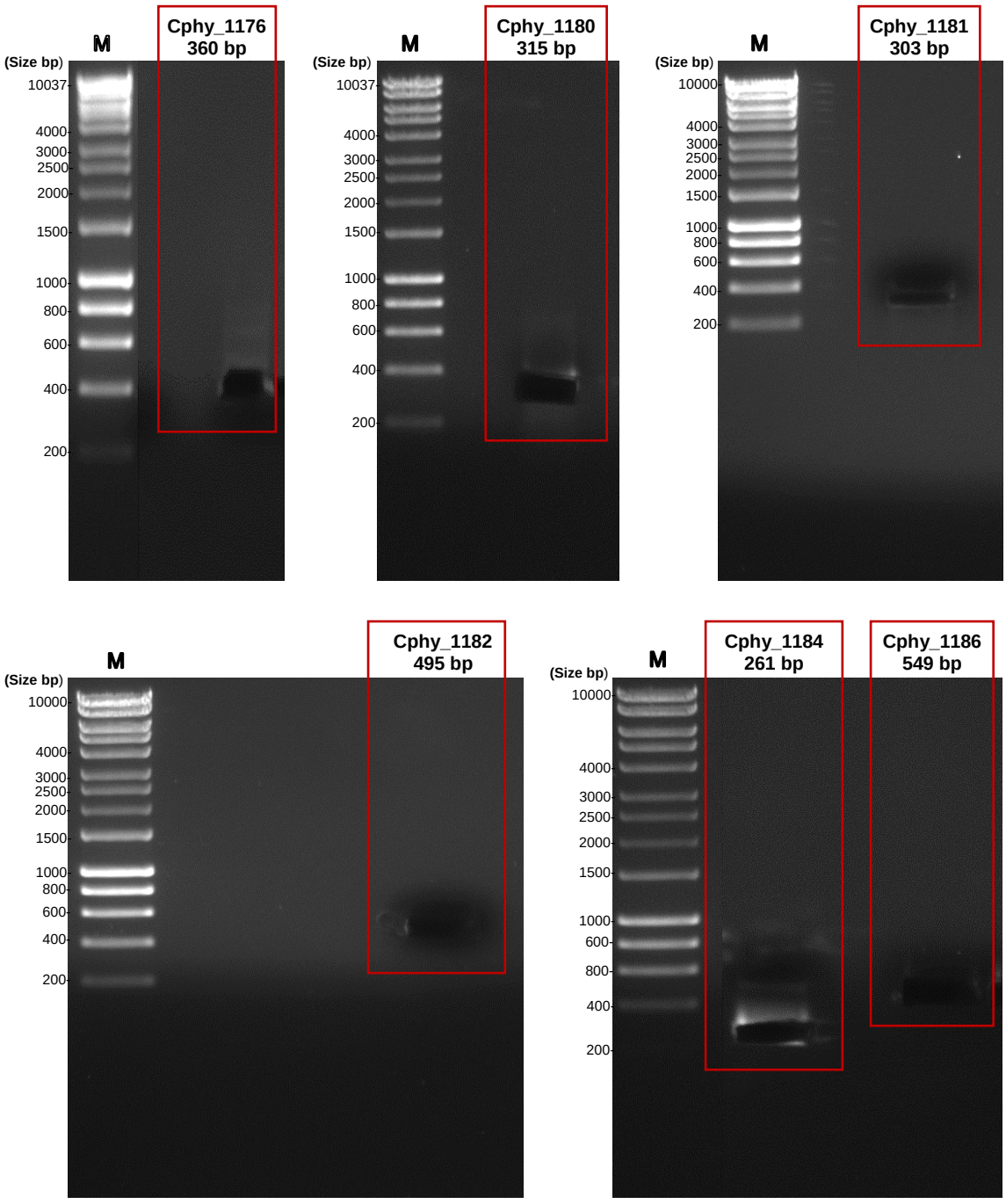


Figure 3.2: Gel electrophoresis of PCR products of the six shell proteins Cphy1176, Cphy1180, Cphy1181, Cphy1182, Cphy1184 and Cphy1186. The photos were taken after cutting the bands. This gel made of 1 % Agarose.

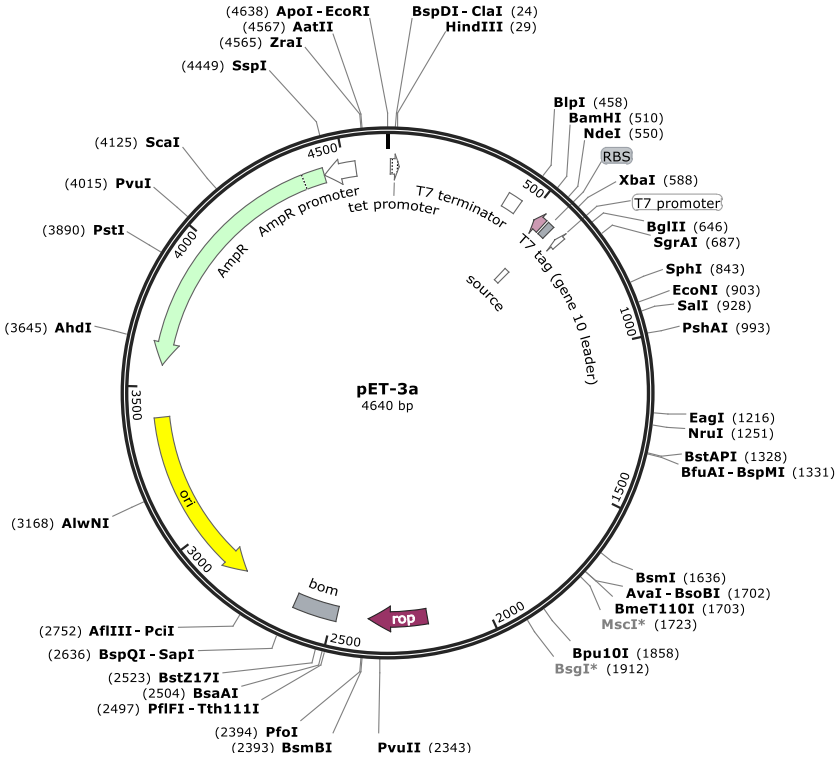


Figure 3.3: pET-3a vector including cloning, expression, and T7 N-terminal Tag region. Cat. No is 69418-3 Novagen.

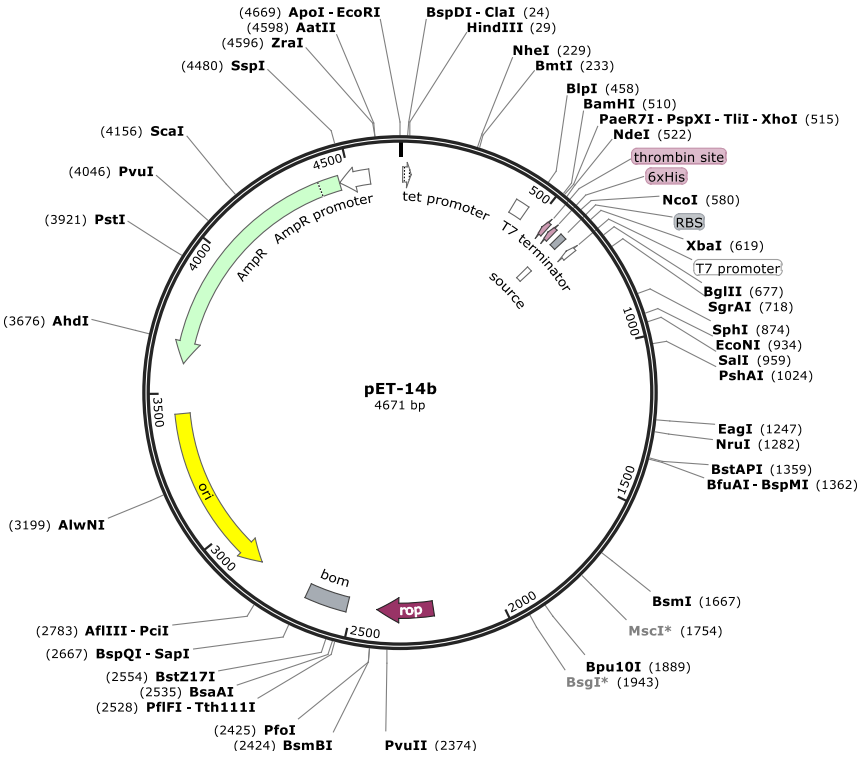


Figure 3.4: pET-14b vector including cloning, expression, and T7 N-terminal Tag, His Tag region. Cat. No is 69660-3 Novagen.

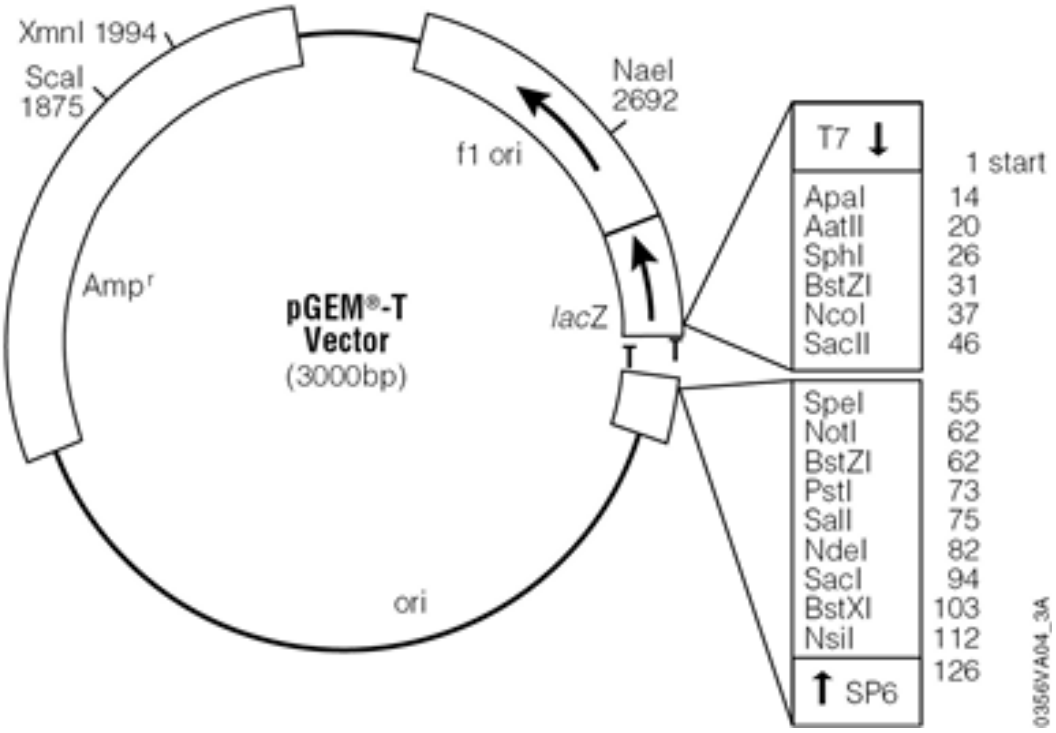


Figure 3.5: The pGEM®-T Vector is derived from the pGEM®-5Zf(+) Vector (GenBank® Accession No. X65308). The pGEM®-T Vector was created by linearizing the pGEM®-5Zf(+) Vector with EcoRV at base 51 and adding a T to both 3´-ends.

3.2.3 Shell protein overproduction, purification and characterisation

The pET14b plasmids express proteins that encode an N-terminal hexa-histidine tag to allow for rapid purification using immobilised metal affinity chromatography (IMAC). Plasmids encoding the six individual shell proteins, were transformed into *E. coli* BL21 Star™ (DE3)pLysS. The transformants were grown at 37°C in LB media with the addition of appropriate antibiotics, ampicillin and chloramphenicol. Cultures were incubated until the OD_{600} reading was approximately 0.6 **(Section 2.2.7)**. At this point protein

overproduction was induced via addition of IPTG prior to the cultures being incubated overnight at 18°C. After overnight induction the cultures were harvested by centrifugation and the protein extracted after sonication of the resuspended bacterial pellet. The soluble protein was purified by IMAC and the fractions containing the appropriate shell protein was then desalted on a PD-10 column. A final purification involving gel-filtration on a fast protein liquid chromatography (FPLC) system generated protein that migrated as a single band when analysed via SDS-PAGE. All the shell proteins, with the exception of Cphy_1176 which was insoluble, were successfully purified via this isolation strategy.

3.2.3.1 Characterisation and solubilisation of Cphy_1176

The *Cphy_1176* gene is 360 bp in length and the encoded protein has a size of 12.9 kDa. In comparison to the well characterised Pdu BMC system, Cphy_1176 aligns most closely with PduU displaying 45 % sequence identity **(Figure 3.6)**. When Cphy_1176 was purified, the majority of the overproduced protein was insoluble and remained in the pellet fraction **(Figure 3.7)**. Different growth conditions (28°C or 37°C) and changes in induction (different IPTG concentrations or no induction), as well as different growth media (LB or AIM (Autoinduction media)) were all tested in order to try to improve the level of soluble protein. However, no significant improvement was realised with these variations **(Table 3.4 and Figure 3.8)**. Nonetheless, purification of the low level of soluble protein was attempted.

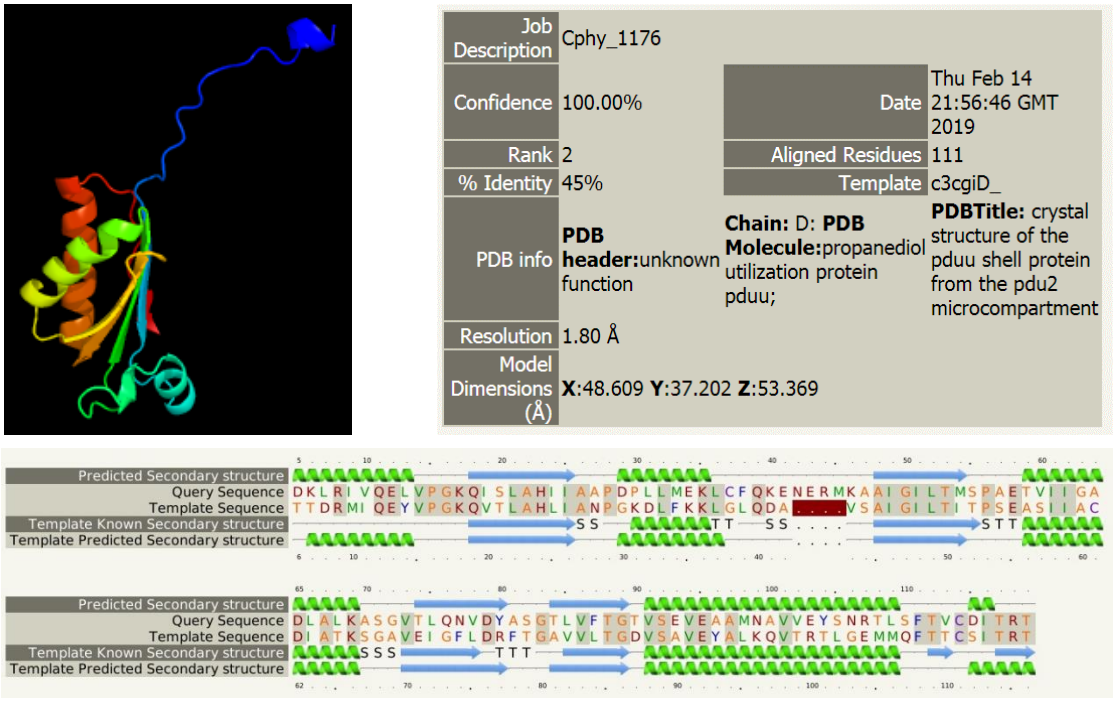


Figure 3.6: Sequence analysis for *C. phytofermentans* Cphy_1176 shell protein compared with PduU. The top left panel displays the modelled 3D structure. The top right panel reports the data concerning the query protein and template structure. The bottom panel highlights the sequence alignment between the two proteins. This Figure was produced using Phyre[2] (Kelley *et al*., 2015).

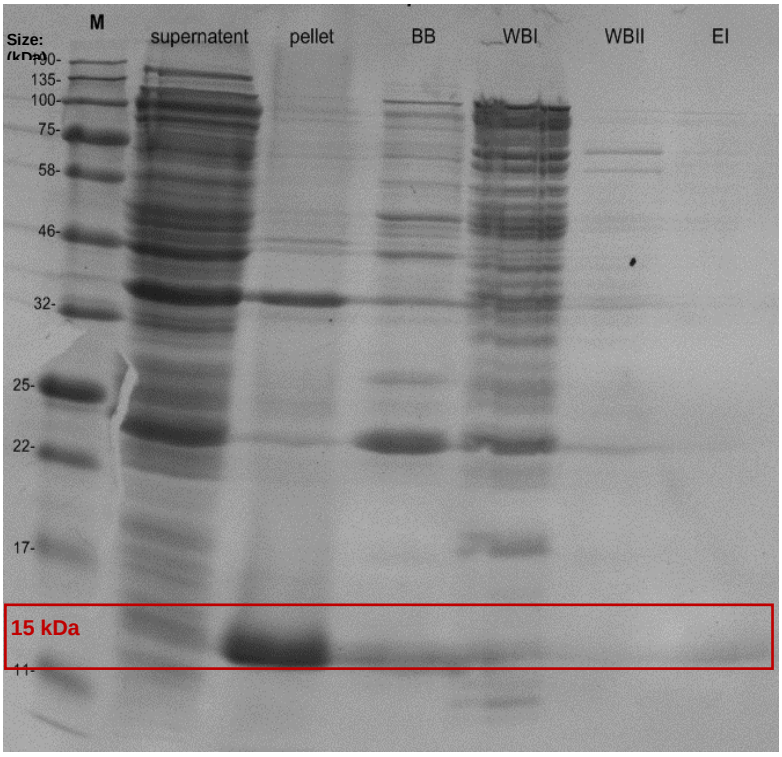


Figure 3.7: 15 % SDS-PAGE gel showing the IMAC purification of Cphy_1176 shell protein. 10 µL of each fraction was loaded. The protein size, including the Hexa-His-Tag, is shaded with red colour. The symbols for the lanes are: M = molecular mass marker, Supernatant = soluble extract, Pellet = insoluble fraction, BB = eluate from binding buffer, WBI = eluate from washing buffer I, WBII = eluate from washing buffer II and EI = eluate from elution buffer.

Table 3.4: Ten different conditions were used to solubilise the shell protein Cphy_1176. Each column is shaded corresponding to the colour of the lane in the SDS-PAGE photo (Fig. 3.8) to be identified easily.

Culture	1	2	3	4	5	6	7	8	9	10
LB/ mL	-	-	100	100	100	100	100	100	100	100
AIM/ mL	100	100	-	-	-	-	-	-	-	-
Tm/ °C	28	37	28	37	28	37	28	37	28	37
Time/ h	24	24	2	3	24	24	3	2	24	24
IPTG concentration	-	-	40	40	-	-	40	40	40	40

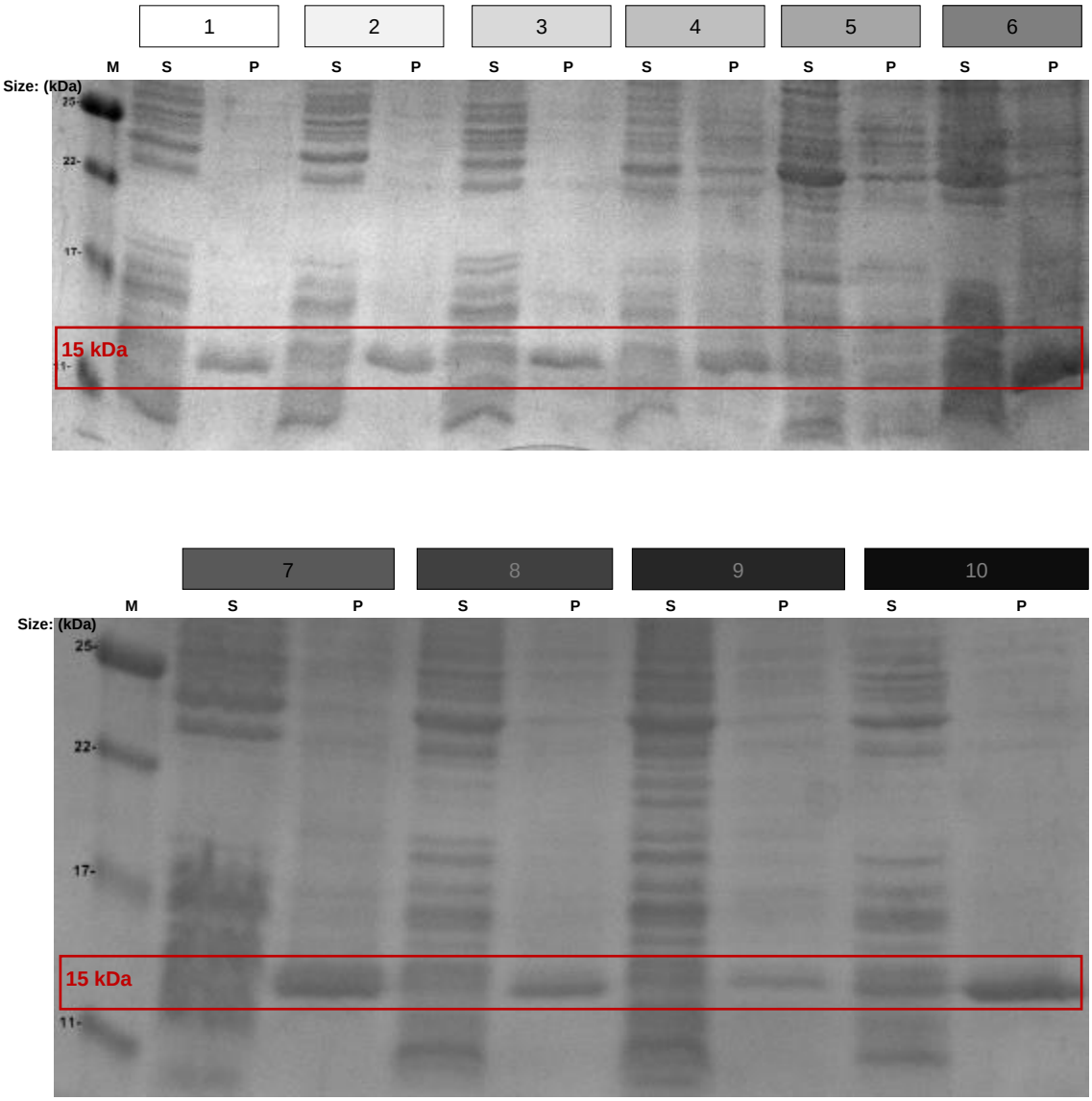


Figure 3.8: 15 % SDS-PAGE gel showing soluble (S) and insoluble (P) fractions of overexpressed Cphy_1176 shell protein. The protein was harvested from 50 mL different cultures. The protein size, including the Hexa-His-Tag, is shaded with red colour.

Some of these variations showed more protein in the pellet than others, such as conditions 6, 7, and 10. However, none of these conditions showed more soluble protein. An alternative purification technique was used to also try and increase the solubility of Cphy_1176. This involved the use of B-PER™ reagent to resuspend the cell pellet instead of binding buffer, as the B-PER™ reagent contains some mild detergents that can help solubilise protein. This approach resulted in purification of a sufficiently soluble protein for analysis **(Figure 3.9)**. When the protein was purified by the standard purification technique was loaded onto an SDS-PAGE gel **(Figure 3.7)**, there was no obvious band in the elution fraction. By contrast, the use of the B-PER™ reagent resulted in the appearance of a faint band after analysis by SDS-PAGE **(Figure 3.9)**. However, the isolated Cphy_1176 protein could not be concentrated in an Amicon® Ultra 10K device (10,000 NMWL) which was

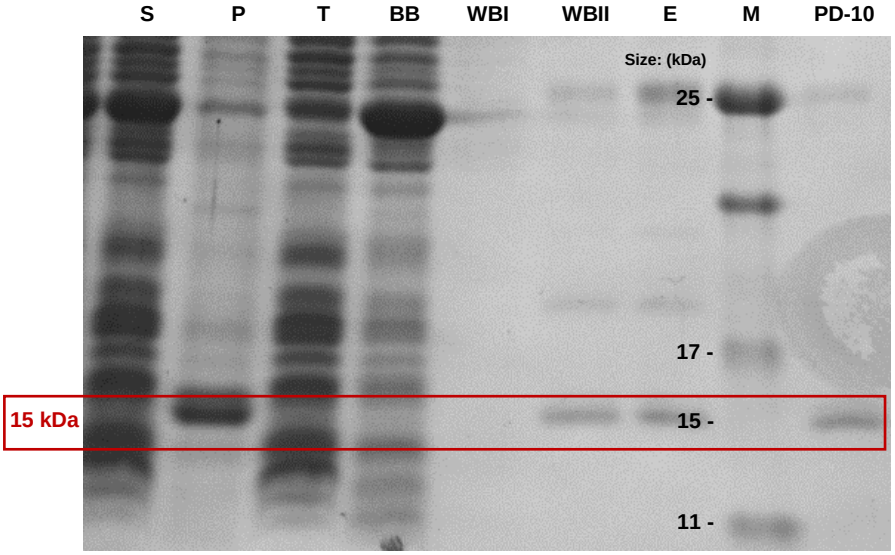


Figure 3.9: 15 % SDS-PAGE gel showing IMAC & PD-10 purification process fractions of soluble (S), insoluble (P), total (T), binding buffer (BB), washing buffer I (WBI), washing buffer II (WBII) and elution (E) of overexpressed Cphy_1176 shell protein. The protein was harvested from a 1000 mL culture. Cells were lysed via B-PER, and sonication. The protein size, including the Hexa-His-Tag, are shaded with red colour.

necessary prior to further FPLC purification. Several attempts were made to purify and concentrate this protein but it always precipitated out during the concentration stage.

In conclusion, the recombinant production of Cphy_1176 (PduU-like) in *E. coli* results in largely insoluble protein which when purified tends to precipitate at higher protein concentrations.

3.2.3.2 Characterisation of Cphy_1180

The Cphy_1180 protein most closely resembles the PduA protein from the Pdu system, displaying a sequence identity of 42 % **(Figure 3.10)**.

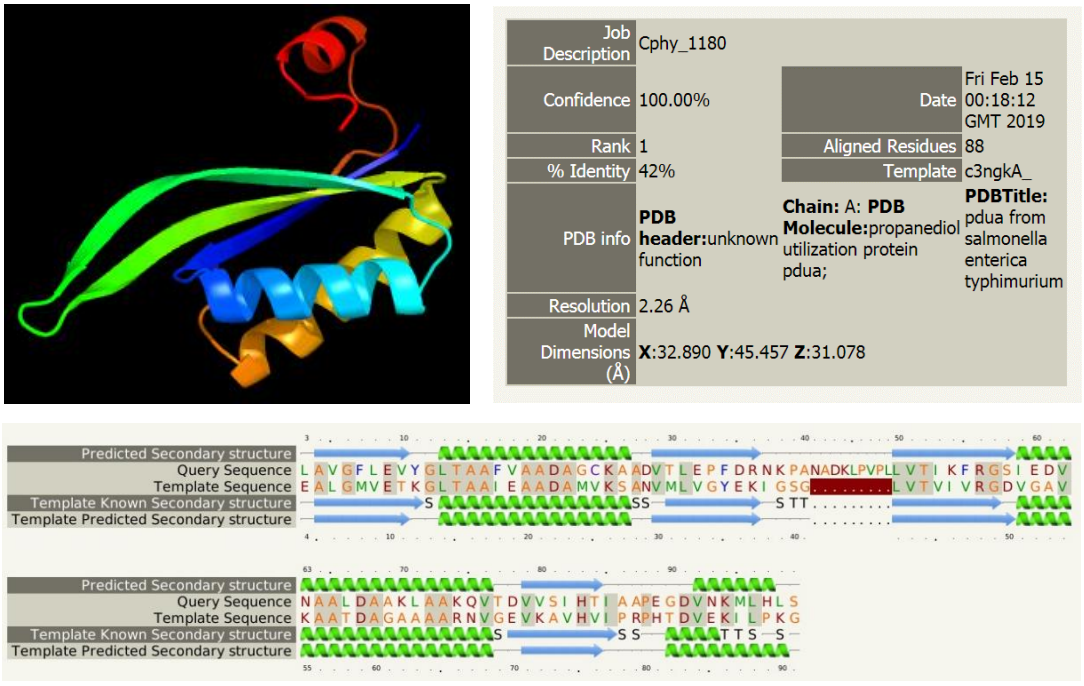


Figure 3.10: Sequence analysis for *C. phytofermentans* Cphy_1180 shell protein compared with PduA. On the top left is the assumed 3D structure. On the top right are some data about template and the query protein. At the bottom is the sequence alignment. This Figure was produced using Phyre[2] (Kelley *et al*., 2015).

Cphy_1180 has a predicted molecular mass of 10.9 kDa. A 1 L culture of *E. coli* overproducing Cphy_1180 was harvested and the cell pellets re-suspended in binding buffer and stored at – 80°C. After gentle thawing, the cells were lysed by sonication as described in **Chapter 2**. The lysate was centrifuged and the supernatant applied to the nickel charged column **(Figure 3.11)**. When recombinantly produced in *E. coli*, a significant amount of the protein was observed to be insoluble. Nonetheless, enough of the protein remained soluble to allow the protein to be purified to homogeneity **(Figure 3.12)**. The PD-10 fraction was concentrated to 1.5 mL and then further purified via FPLC using a G200 column.

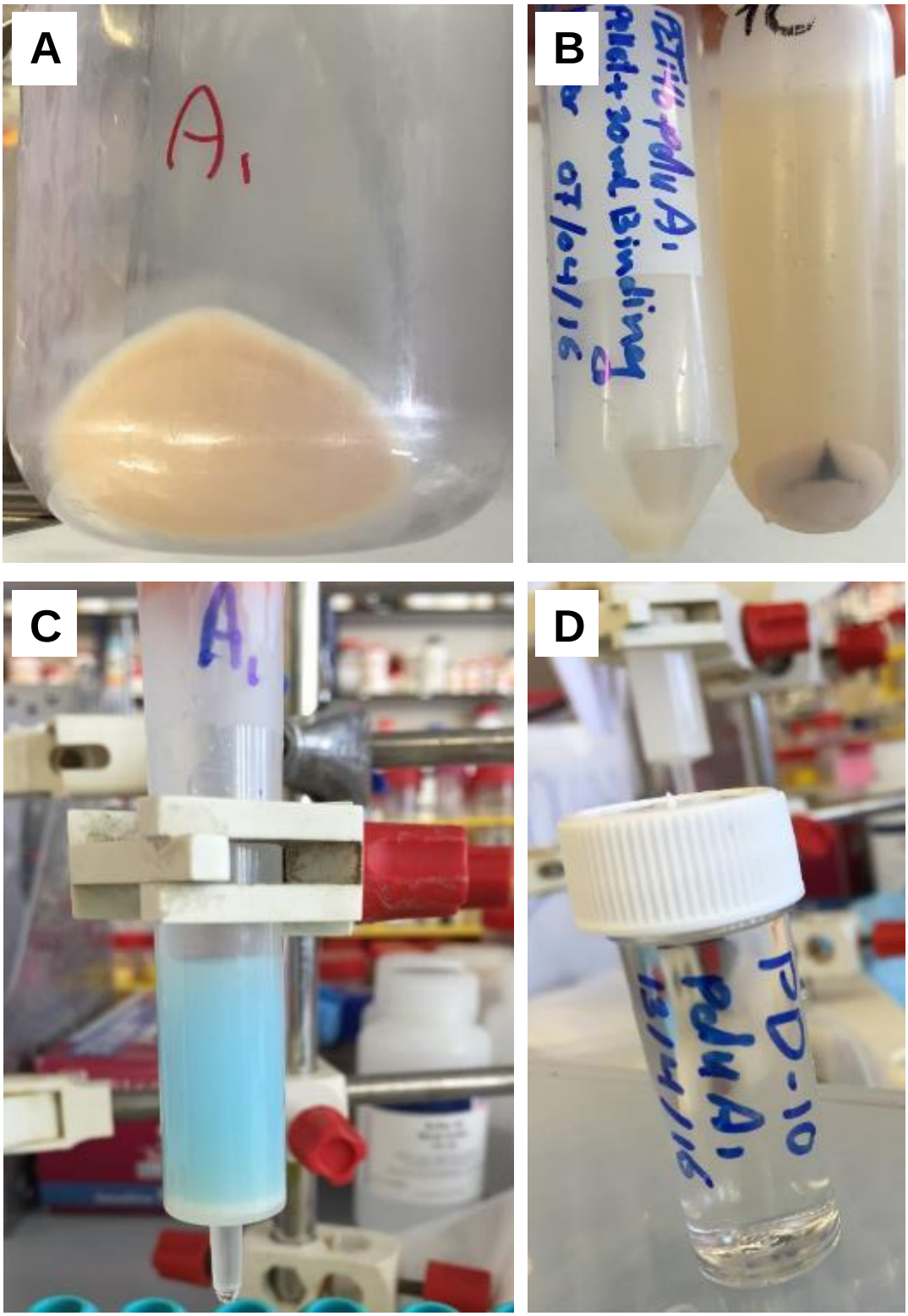


Figure 3.11: IMAC purification of Cphy_1180. (A) photo of the harvested cells before partly lysed with binding buffer (B) photo of the lysed centrifuged cells including the insoluble protein at the bottom of the tube. (C) photo for the applied protein at the last stage of purification after washing and before elution. (D) photo for purified Cphy_1180 after desalting.

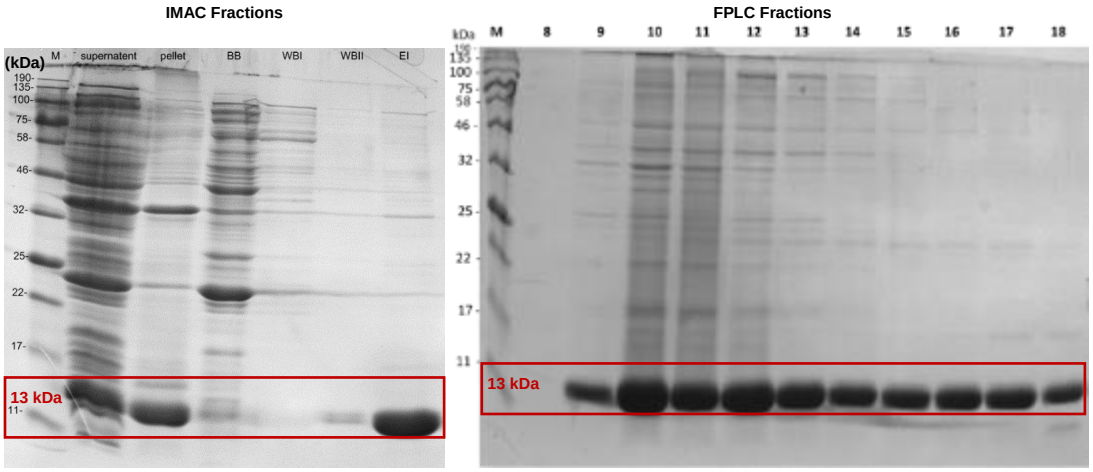


Figure 3.12: 15 % SDS-PAGE gel of fractions from Cphy_1180 shell protein purification. 10 µL of each fraction was loaded. The protein size, including the Hexa-His-Tag, are shaded with red colour. On the left are IMAC purification fractions and on the right are FPLC fractions. Labels: Soluble (Supernatant), insoluble (Pellet), binding buffer (BB), washing buffer I (WBI), washing buffer II (WBII), and elution (EI).

By gel filtration chromatography the Cphy_1180 was found to elute in two separate peaks, after a volume of 18.58 mL as well as at 30.68 mL **(Figure 3.13)**. The presence of the protein in these peaks was verified by SDS-PAGE **(Figure 3.12)**. The column had been previously calibrated with a range of protein standards from a Gel Filtration markers kit **(Sigma)**. Plotting a standard curve of elution volumes for proteins of known molecular mass allowed the estimation of the size of the proteins in this experiment. The first peak observed at an elution volume of 18.58 mL equates to a protein in the mega-Dalton range, suggesting this was aggregated protein. The second peak at 30.68 mL equates to a protein of approximately 63 kDa, representing a potential hexamer **(Table 3.5)**.

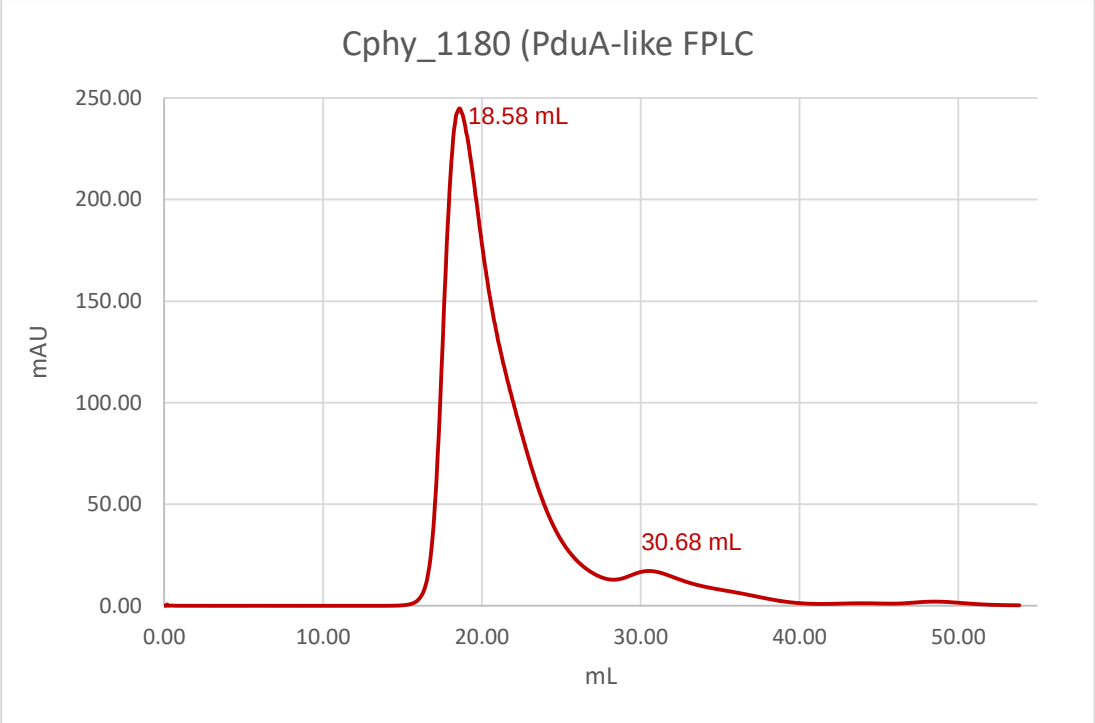


Figure 3.13: Aerobic Purification and gel filtration of Hexa-His tagged Cphy_1180 protein (pduA-like). This curve shows the FPLC trace (Sephadex G-200) Cphy_1180 protein in 20 mM Tris-HCl (pH 8.0), with 100 mM NaCl.

Table 3.5: Peaks and calibration values of Cphy_1180 shell protein.

Cphy1180-6-His	Size/ kDa	Peak volme/ mL	Peak size/ kDa	Number of subunits
First peak	13	18.58	15333	1207
Second peak		30.68	63	6

This result is consistent with some of the protein forming large aggregates whereas the rest exists as a hexamer as is found in the shell-protein tiles that make up the shell of the BMC.

3.2.3.3 Characterisation of Cphy_1181

The Cphy_1181 shell protein is predicted to be 10.7 kDa in size. Cell pellets from two 1 L cultures of *E. coli* overproducing Cphy_1181 were re-

suspended in binding buffer and stored at – 80°C. After gentle thawing, the cells were lysed by sonication as described in **Chapter 2**. The lysate was centrifuged and the supernatant applied to the nickel charged column **(Figure 3.14)**. The column was next washed with buffer containing three different imidazole concentrations,10 mM, 50 mM and 100 mM **(Section 2.1.7.1)**. Protein fractions were collected in buffer containing 400 mM imidazole, followed by PD-10 desalting in order to remove the imidazole.

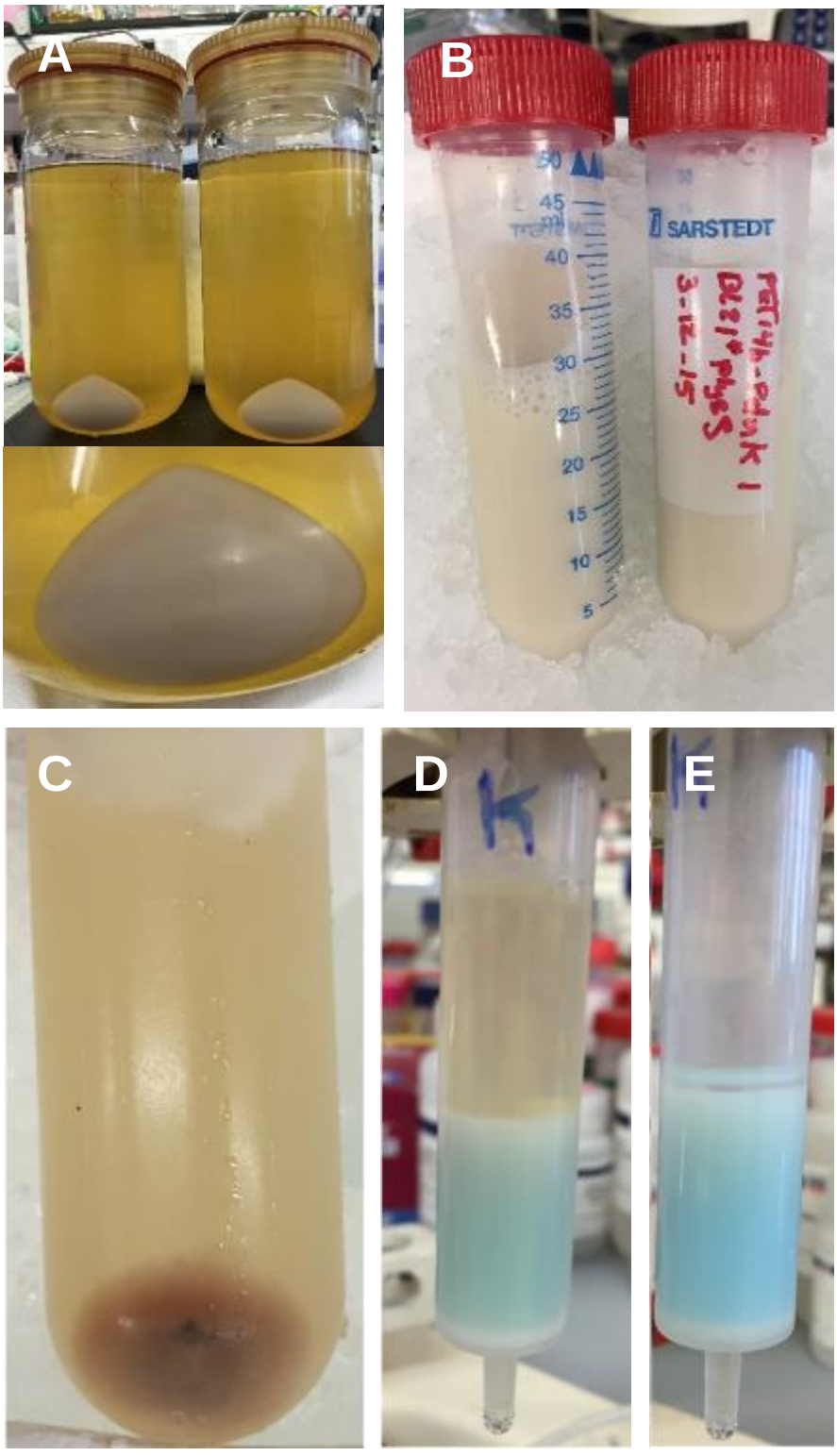


Figure 3.14: IMAC purification of Cphy_1181. (A) photo of the harvested cells (B) photo of the resuspended cells before centrifugation. (C) photo of the lysed centrifuged cells including the insoluble protein at the bottom of the tube. (D) photo for the Cphy_1181 soluble protein loaded into the Ni column. (E) photo for the applied protein at the last stage of purification after washing and before elution.

Samples (10 µL) were taken from each fraction for SDS-PAGE **(Figure 3.15)**. However, when Cphy_1181 protein from the frozen cell pellets was purified using the fresh buffers pH 8.0 **(Section 2.1.7.1)** then treated with

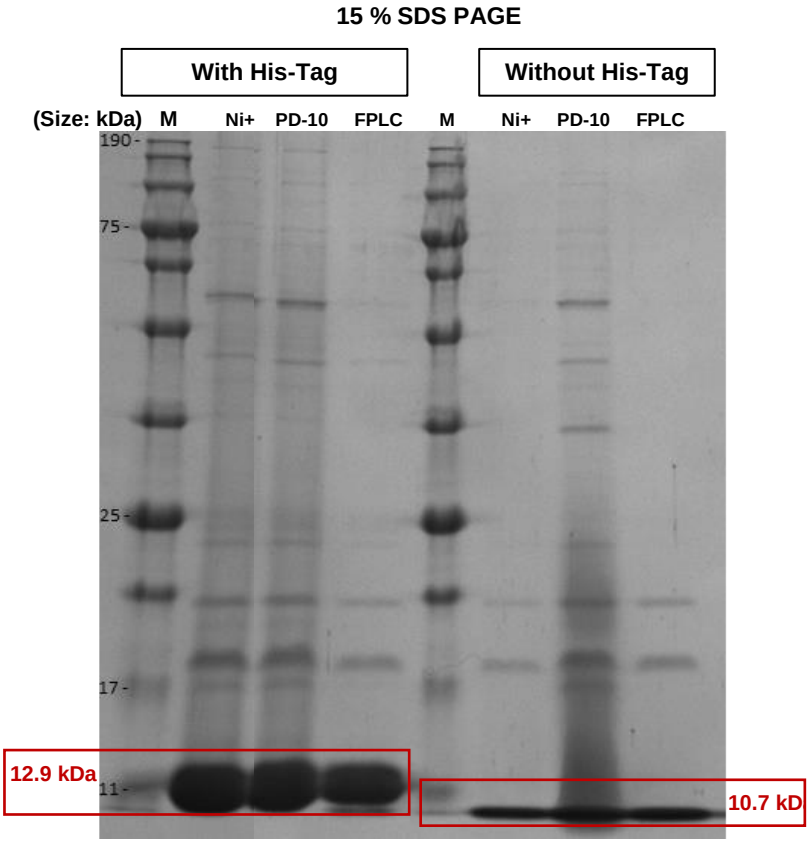


Figure 3.15: 15 % SDS-PAGE gel photo of Cphy_1181 purification. This photo displays the fractions of all the three purification techniques IMAC, PD-10, and FPLC with 6His-Tag and without it. The protein sizes are shaded with red colour.

Bio-rad protein reagent in order to select the fractions with highest protein concentration, a bright turquoise colour was displayed instead of the blue as all the other shell proteins showed **(Figure 3.16)**.

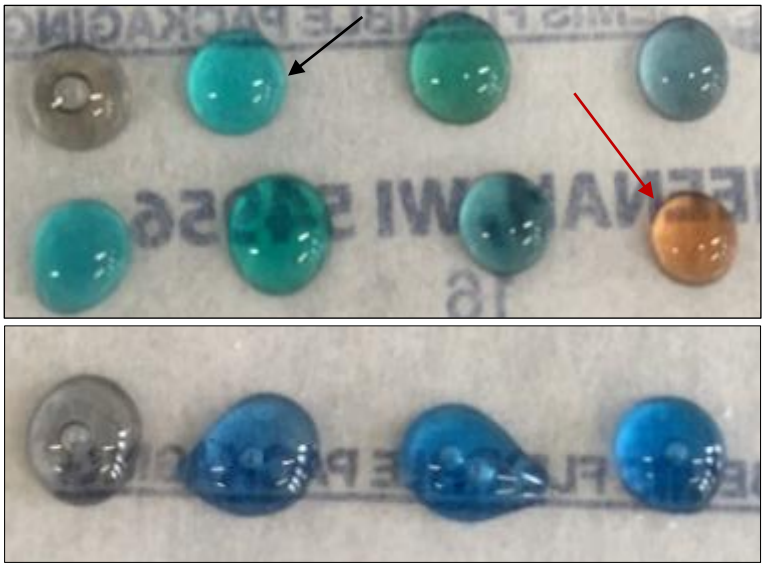


Figure 3.16: Purified proteins treated by Bio-rad protein reagent. On the top the original colour of the Bio-rad reagent is brown as shown by the red arrow and the colour of the Bio-rad treated Cphy_1181 protein is bright turquoise as shown by the black arrow. On the bottom the Bio-rad treated a different shell protein is blue.

Following IMAC purification, the Cphy_1181 protein was soluble but once the imidazole was removed using a PD-10 column the isolated protein precipitated out. Alternative buffers at a range of pH and salt concentrations were tested but no improvement to the solubility of the protein was observed. It was suggested that the insolubility may be due to the presence of the HexaHis-tag on the protein. To investigate this the protein was treated with a small amount of thrombin, since the HexaHis-tag is attached to the protein via a thrombin cleavage site. Surprisingly, just seconds after the thrombin addition, the protein appeared to go back into solution **(Figure 3.17)**. Hence, protein solubility in this case was significantly improved by removal of the HexaHis-tag. Analysis by SDS-PAGE indicated that purified Cphy_1181 protein with HexaHis-tag was not very stable and was degraded after 24 h post purification (data not shown). Whilst most of the Cphy shell proteins

concentrated easily in the spin-concentrators, from 3.5 mL to 1.5 mL, within 15 mins, Cphy_1181 took more than an hour to concentrate from 3.5 mL to 2 mL.

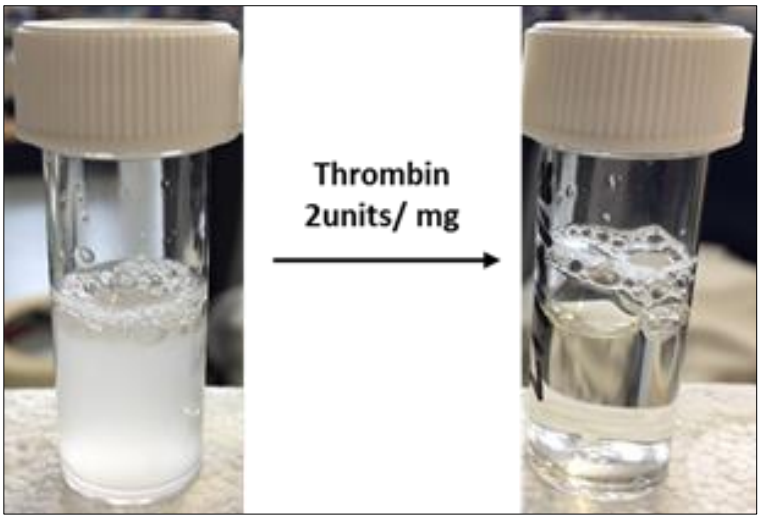


Figure 3.17: Photo of the Cphy_1181 (PduK-like) transformation from precipitated to soluble protein. This transformation occurred immediately after adding 2 units/ mg of Thrombin in room temperature.

Using a G200 column attached to an FPLC system, Cphy_1181 was found to elute at volume of 15 mL **(Figure 3.18)**. The Cphy_1181-HexaHis peak was observed to elute at a volume of 14.94 mL, equating approximately to a hexamer. The Cphy_1181 that had its HexaHis tag removed was found to elute at a similar volume, 15.30 mL, which also equates approximately to a hexamer **(Table 3.6)**. SDS-PAGE of the eluted peaks indicates the main protein is Cphy_1181 11 kDa with no contamination present **(Figure 3.15)**.

These results are consistent with the protein existing as a hexamer, which of course is the predicted outcome.

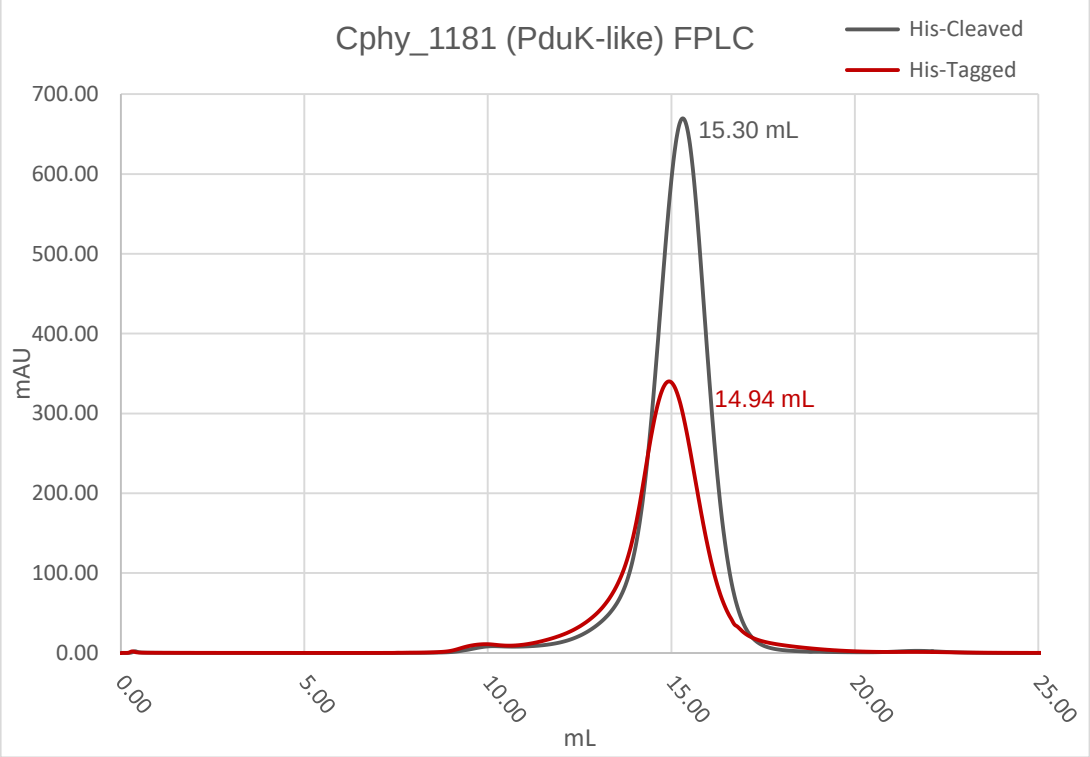


Figure 3.18: Further purification of Cphy_1181 via gel filtration chromatography. G200 elution profile of both proteins 6His-tagged and the 6His-cleaved displaying peaks volume around 15 mL.

Table 3.6: Peaks and calibration values of Cphy_1181 shell protein with and without Hexa-His tag.

Proteins	Size/ kDa	Peak volume/ mL	Peak size/ kDa	Likely number of subunits
Cphy_1181-Hexa-His	12.9	14.94	103	6
Cphy_1181	10.7	15.30	88	6

3.2.3.4 Characterisation of Cphy_1182

The Cphy_1182 shell protein has a predicted size of 17 kDa. This is another PduA-like protein but in this case, it has an N-terminal extension of approximately 75 amino acids (MDDKLDKKSNSTRKAEITGEGKKSRSTSKVSLTGQSIDAVSGSNISMEVM KANKNINEEVTVDTTVLKEEKIMAQ). This N-terminal extension does not display similarity with any other protein of known function. Cell pellets from two 1 L cultures of *E. coli* overproducing Cphy_1182 had previously been re-suspended in binding buffer and stored at – 80°C. After gentle thawing, the cells were lysed by sonication as described in **Chapter 2**. The lysate was centrifuged and applied to a nickel charged column **(Figure 3.19)**. The protein was washed with buffer containing three different imidazole concentrations, 10 mM, 50 mM and 100 mM. Cphy_1182-containing fractions were eluted in buffer containing 400 mM imidazole **(Section 2.1.7.1)**. Samples from the purification were collected and analysed by SDS-PAGE **(Figure 3.20)**. There are no Trp, Tyr or Cys residues present in Cphy_1182 so the protein does not absorb strongly at 280 nm by UV spectrophotometry. The concentration of the protein could therefore not be quickly calculated. Thrombin was added in order to cleave the HexaHis-Tag. The protein was incubated with thrombin at room temperature on the rotator for 4 to 5 h. After cleavage, the protein solution was desalted and then concentrated to 1.5 mL.

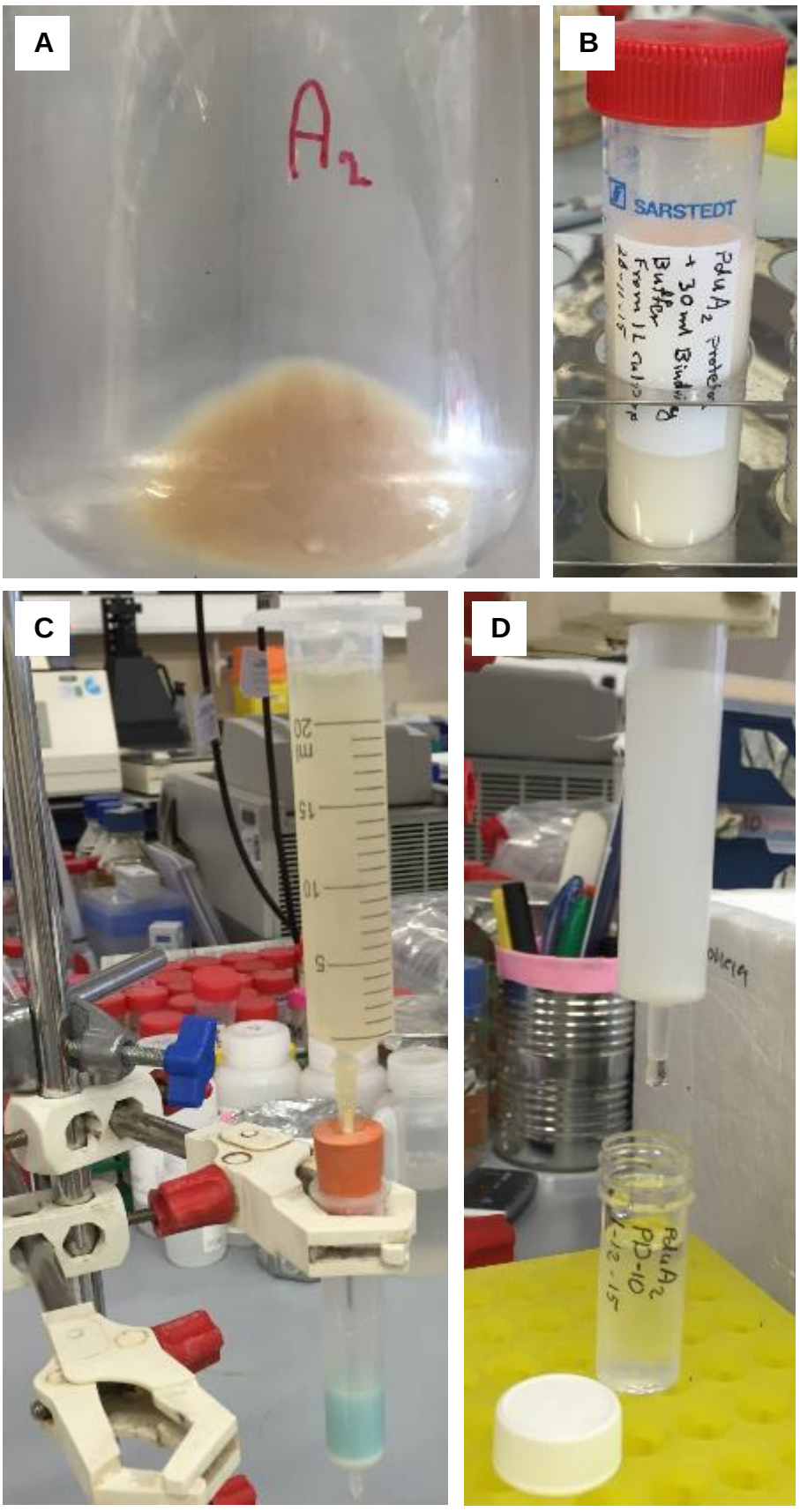


Figure 3.19: IMAC purification of Cphy_1182. (A) photo of the harvested cells (B) photo of the lysed cells before centrifugation. (C) photo for the Cphy_1182 soluble protein loaded into the Ni column. (D) photo of purified Cphy_1182 during desalting.

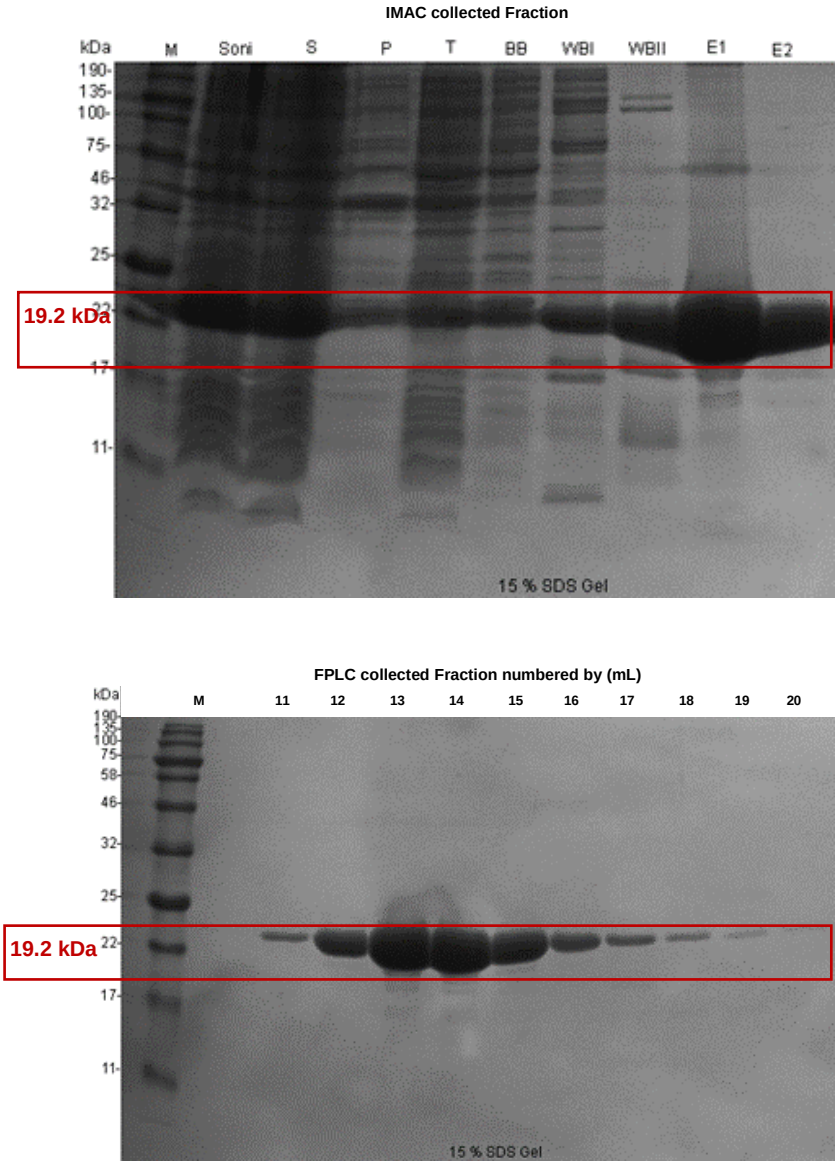


Figure 3.20: 15 % SDS-PAGE gel of Cphy_1182 shell protein purification. 10 µL of each fraction was loaded. The protein size, including the Hexa-His-Tag, is shaded with red colour. On the top are IMAC purification fractions and on the bottom are FPLC fractions by mL. Labels: sonication (Soni), Soluble (S), insoluble (P), total (T) binding buffer (BB), washing buffer I (WBI), washing buffer II (WBII), elutions (E1) and (E2). FPLC photo was taken before cleaving the Hexa-His-Tag.

Using a G200 column, Cphy_1182 eluted at a volume of 13 mL and 19 mL **(Figure 3.21)**. The first peak observed at an elution volume of 13.31 mL equates to a dodecamer (12-mer). Interestingly, some shell proteins are

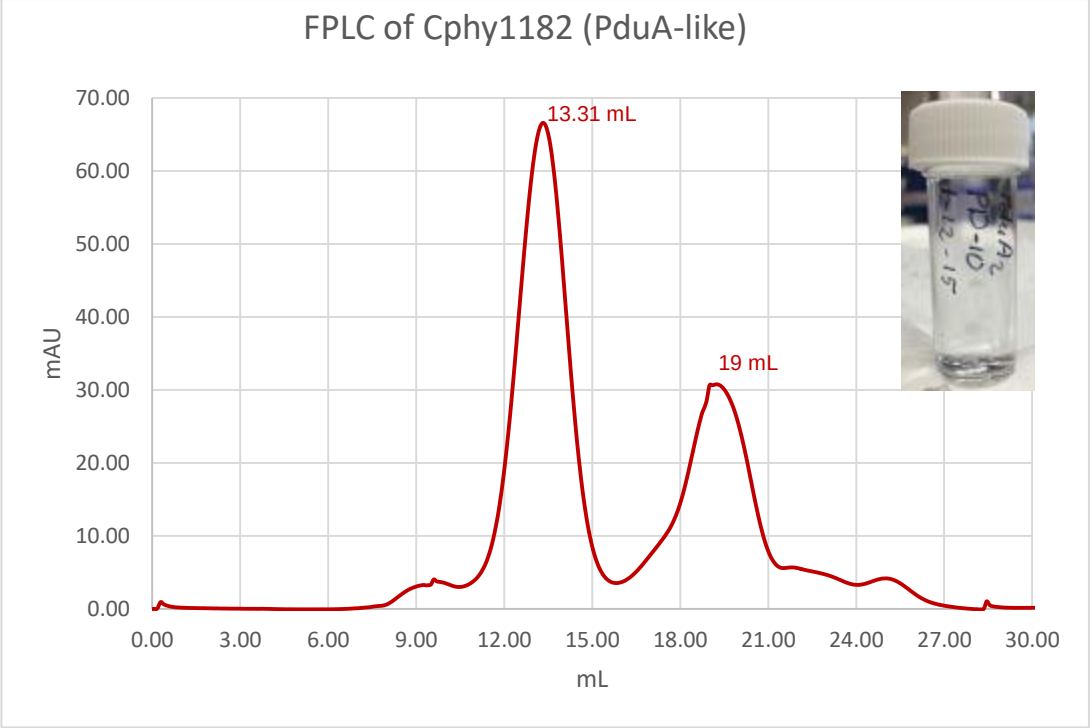


Figure 3.21: Cphy_1182 gel filtration chromatography curve. G200 elution profile of the protein displaying peak volumes around 13 and 19 mL. On the top right is a purified sample of Cphy_1182.

known to form double hexameric-ring structures, which are thought to form a locked pore system (Chun *et al.*, 2014). The second peak at 19 mL equates to the size of a monomer **(Table 3.7).** SDS-PAGE of the eluted peaks indicates the main protein is Cphy_1182 with no contamination present **(Figure 3.20)**. Further information on the characterisation of Cphy_1182 shell protein and its structure is presented in **Chapter 5**.

Table 3.7: Peaks and calibration values of Cphy_1182 shell protein.

Peaks	Size/ kDa	Peak volme/ mL	Peak size/ kDa	Number of subunits
First peak	17	13.31	212	12
Second peak		19	17	1

3.2.3.5 Characterisation of Cphy_1184

Cphy_1184 gene was ligated into pET14b (HexaHis-tag) and the protein product overproduced in *E. coli* **(Section 2.2.7)**. The protein was purified by a combination of IMAC, PD-10, and FPLC chromatography as described for the previous shell proteins. The shell protein Cphy_1184 (PduN-like) **(Figure 3.22)** was observed to have some interesting characteristic. Protein Cphy_1184 displayed a yellow colour at the top of the Ni^{2+} column prior to elution **(Figure 3.23)**.

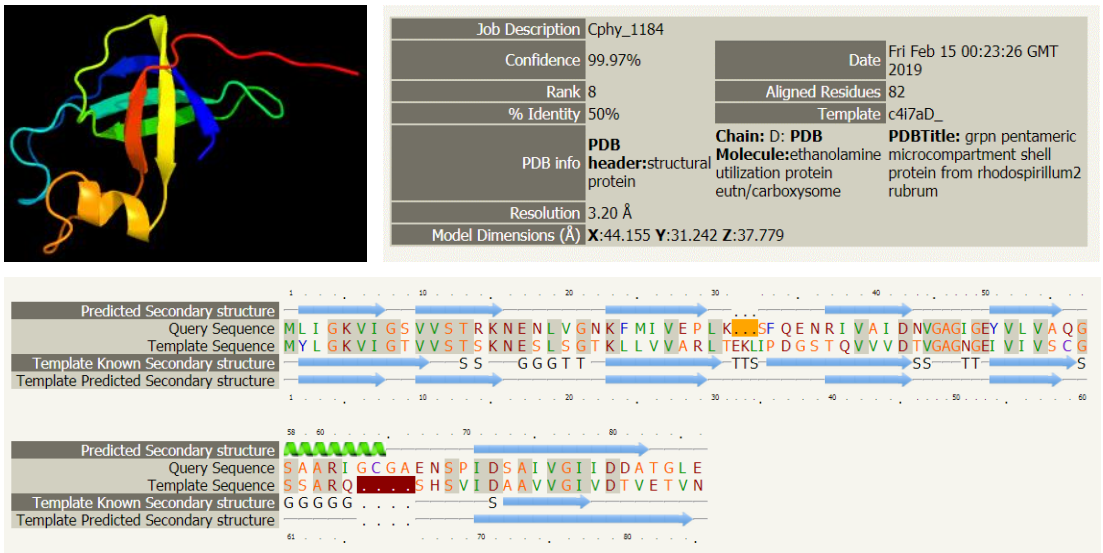


Figure 3.22: Sequence analysis for *C. phytofermentans* Cphy_1184 shell protein compared with EutN. On the top left is the predicted 3D structure. On the top right are some data about Template (EutN) and the query protein. At the bottom is the sequence alignment of Cphy_1184 with EutN. This Figure was produced using Phyre^2 (Kelley *et al*., 2015).

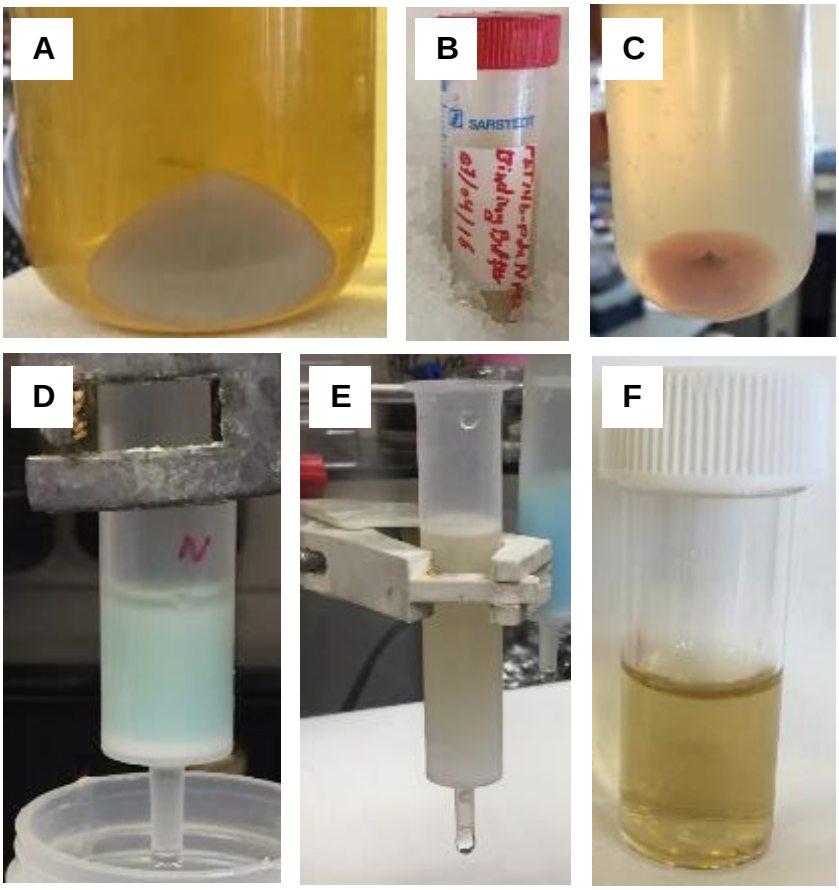


Figure 3.23: IMAC and PD-10 purification of Cphy_1184 protein. (A) a picture of harvested cells. (B) Harvested cells resuspended in binding buffer. (C) Photo of the centrifuged lysate including the insoluble protein at the bottom of the tube. (D) Photo for the Cphy_1184 soluble protein loaded into the Ni+ column at the last stage of purification after washing and before elution. (E) Photo of purified Cphy_1184 during desalting. (F) Photo of purified protein after desalting and concentration.

The band was isolated and samples of each fraction were collected for SDS-PAGE **(Figure 3.24)**. Previously, research had shown that some shell proteins contain Fe-S centres, generating a brown colour, including PduT. However, no colour has ever been previously reported for a PduN-like protein.

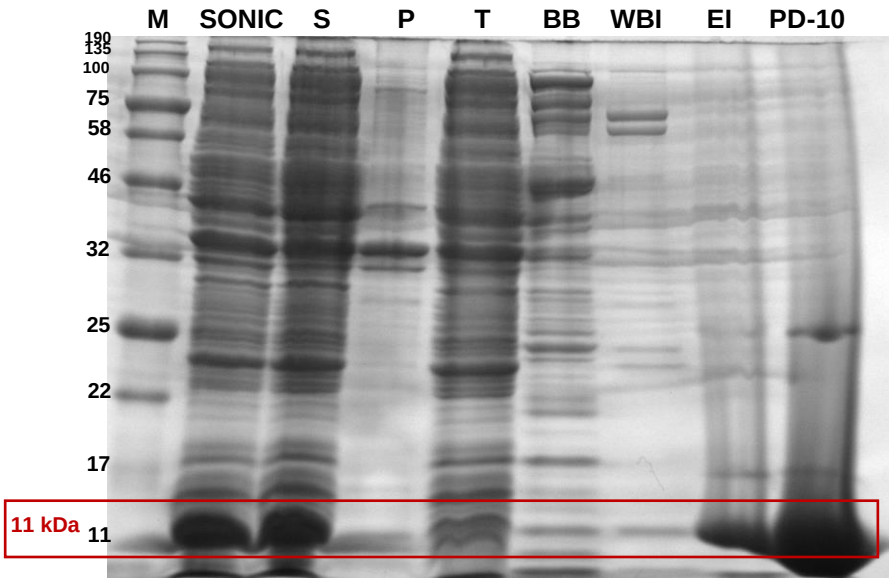


Figure 3.24: 15 % SDS-PAGE gel of Cphy_1184 protein IMAC as well as PD-10 purification processes. 10 µL of each fraction was loaded. The picture is showing fractions of Cphy_1184 proteins and its size including the Hexa-His-Tag, are shaded with red colour. Labels: sonication (SONIC), Soluble (S), insoluble (P), total (T) binding buffer (BB), washing buffer (WBI) and elution (EI).

UV-visible spectra of Cphy_1184 protein was taken and is shown in **Figure 3.25**. Using a G200 column the protein Cphy_1184 eluted at volumes of 13.87 mL **(Figure 3.26)**. SDS-PAGE analysis of the eluted peak showed that the main protein is Cphy_1184 with no contamination present **(Figure 3.26)**. The Cphy_1184 peak observed at an elution volume of 13.87 mL equates to a protein of approximately 15 time the size of Cphy_1184 **(Table 3.8)**.

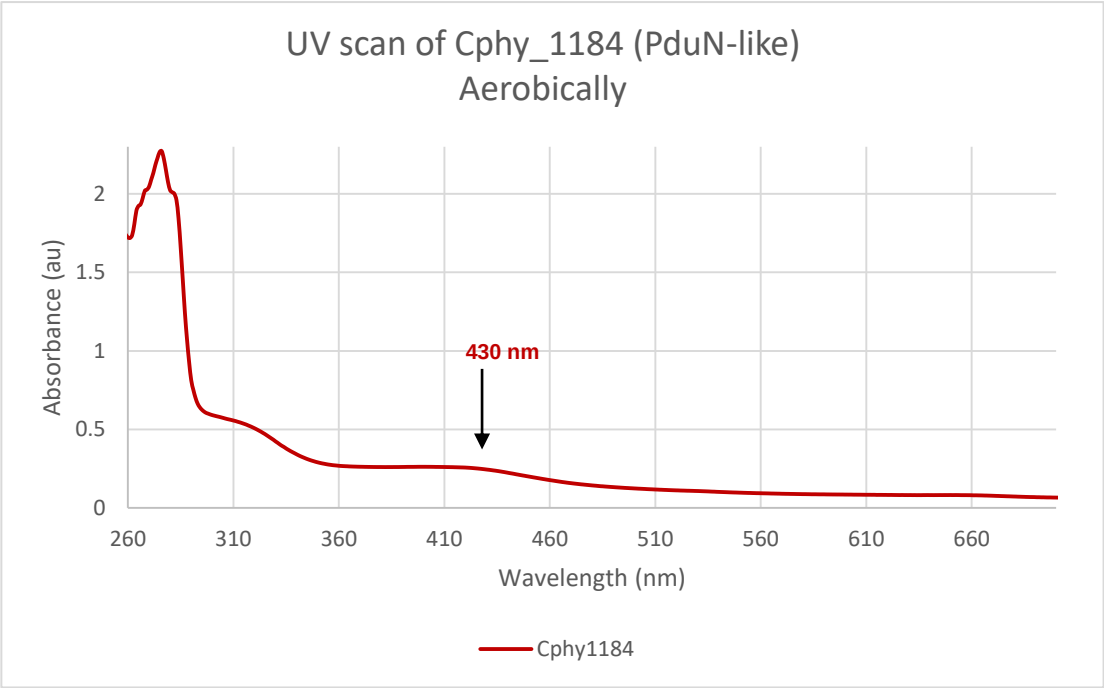


Figure 3.25: UV-visible spectrum of aerobically purified Cphy_1184 shell protein in 20 mM Tris-HCl (pH 8.0), with 100 mM NaCl.

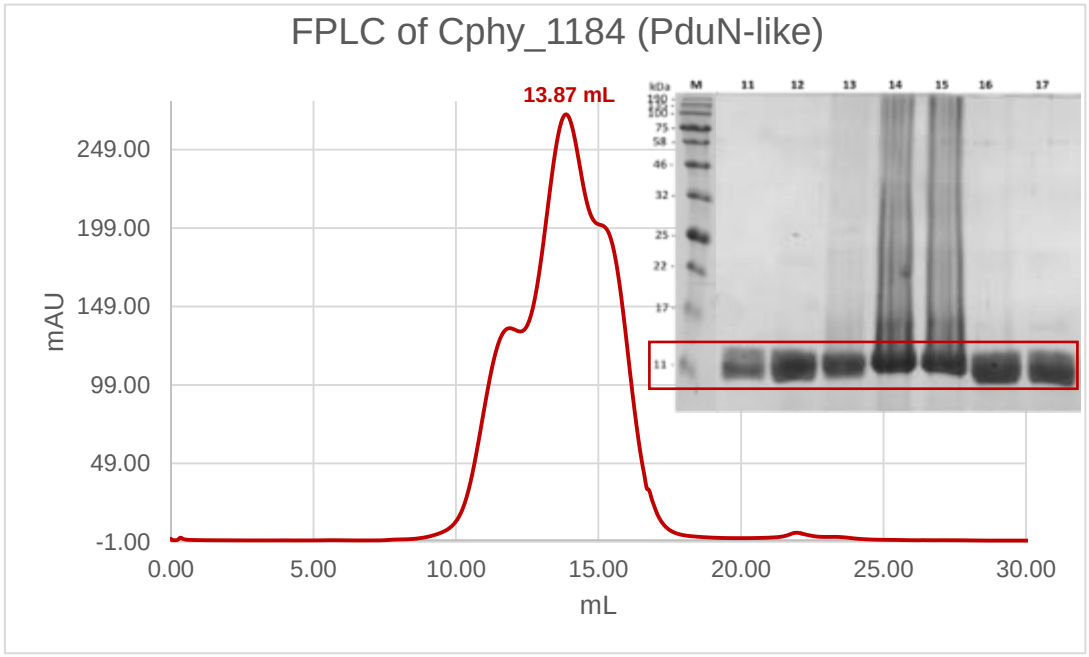


Figure 3.26: Further purification of Cphy_1184 by gel filtration chromatography. A G200 elution profile of the shell protein Cphy_1184 with a peak volume at 13.87 mL with the insert showing an SDS-PAGE gel of the fractions at 11, 12, 13, 14, 15, 16 and 17 mL. On SDS-PAGE the proteins size are shaded with red colour.

Table 3.8: Peaks and calibration values of Cphy_1184 shell protein. The protein size of Cphy_1184 is including the Hexa-His-Tag as it wasn't able to be cleaved.

Protein	Size/ kDa	Peak volume/ mL	Peak size/ kDa	Number of subunits
Cphy_1184	11	13.87	165	15

3.2.3.6 Characterisation of Cphy_1186

Cphy_1186 gene was ligated into pET14b (Hexa-His-tag) and the protein product overproduced in *E. coli* **(Section 2.2.7)**. Cphy_1186 was purified by a combination of IMAC, PD-10, and FPLC chromatography as described for the previous shell proteins. The shell protein Cphy_1186 (PduT-like) was observed to have some interesting characteristic. Protein Cphy_1186 displayed a brown colour at the top of the Ni^{2+} column prior to elution **(Figure 3.27)**. The band was isolated and samples of each fraction were collected for SDS-PAGE **(Figure 3.28)**.

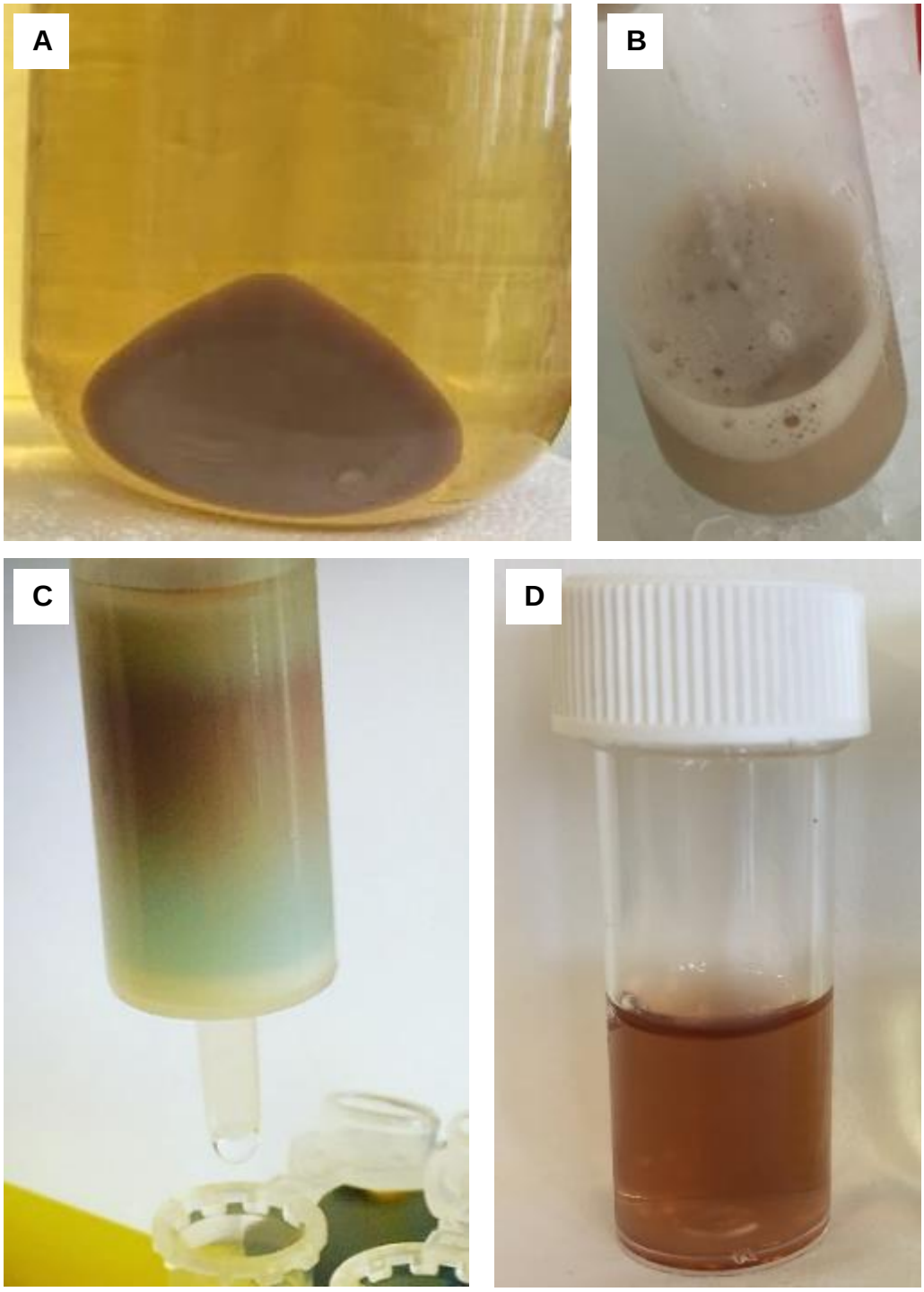


Figure 3.27: IMAC and PD-10 purification of Cphy_1186 protein. (A) a picture of harvested cells. (B) a picture of lysed cells after sonication. (C) a picture of nickel column before elution showing the brown band of the protein in the top of the column. (D) picture of the purified desalted protein showing the brown colour.

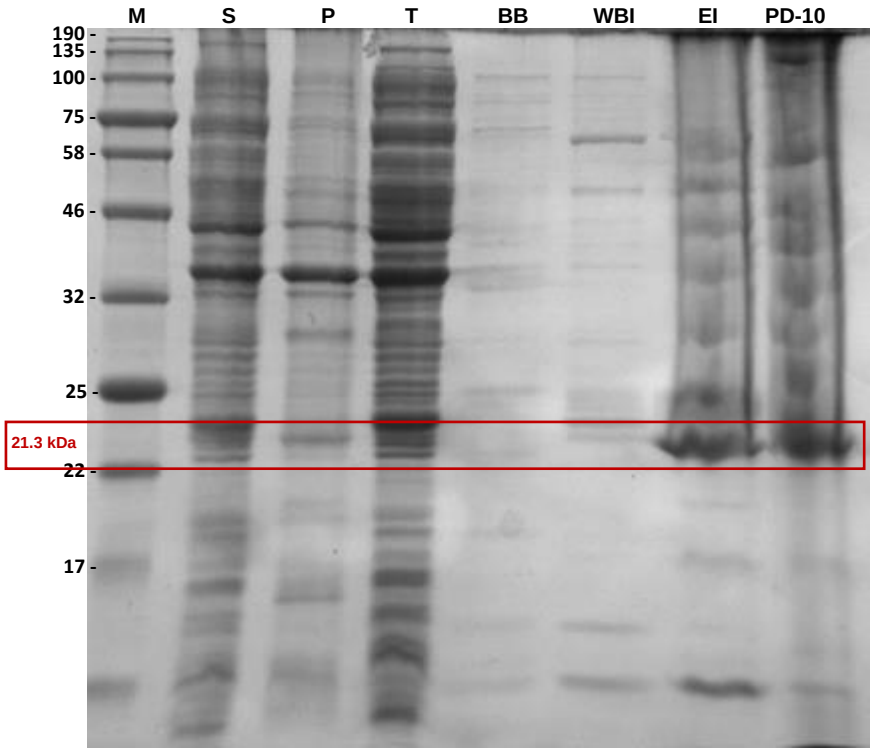


Figure 3.28: 15 % SDS-PAGE gel of Cphy_1186 protein IMAC as well as PD-10 purification processes. 10 µL of each fraction was loaded. The photo is showing fractions of Cphy_1186 proteins and its size is 21.3 kDa including the Hexa-His-Tag. Labels: Soluble (S), insoluble (P), total (T) binding buffer (BB), washing buffer (WBI) and elution (EI). The protein sizes are shaded with red colour.

Previously, research had shown that some shell proteins contain Fe-S centres, generating a brown colour, including PduT. UV-visible spectra of Cphy_1186 was taken and is shown in **Figure 3.29**. The spectra suggest the presence of an iron sulphur cluster [Fe-S] with broad peak between 400 - 430 nm **(Figure 3.29)**. The spectra were enhanced by purification of the protein under anaerobic conditions, which helps with the stability of Fe-S centres. Further characterisation of this protein and EPR analysis of the protein is reported in **Chapter 5**. Using a G200 column the proteins Cphy_1186 eluted at volume of 16 mL **(Figure 3.30)**. SDS-PAGE analysis of the eluted peak showed that the main protein was Cphy_1186 with no contamination present **(Figure 3.30)**.

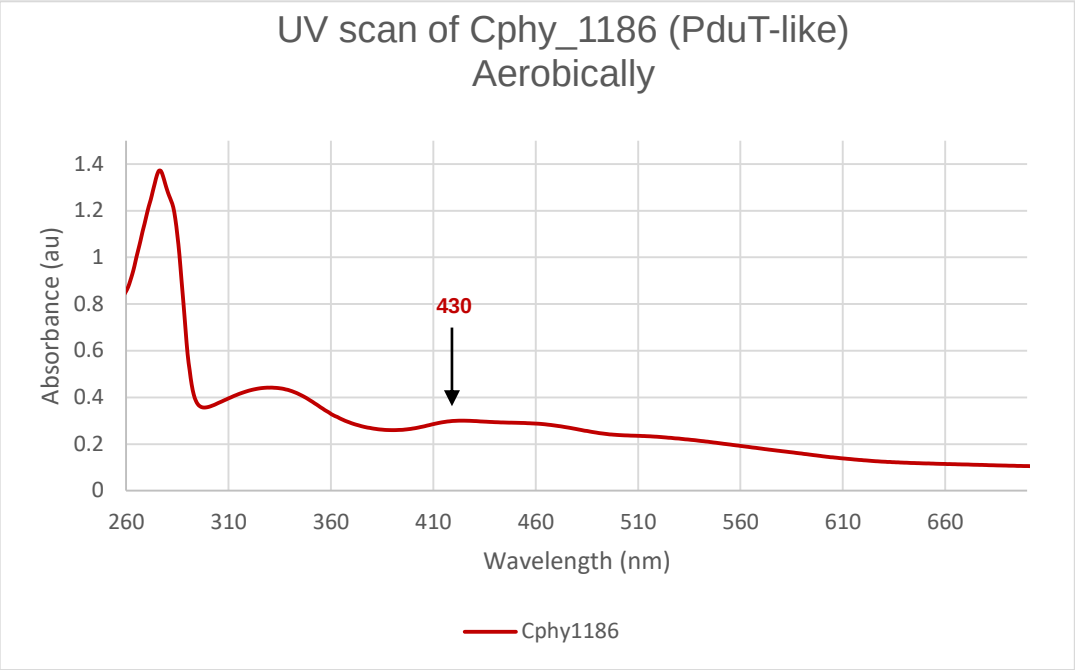


Figure 3.29: UV-visible spectrum of aerobically purified Cphy_1186 shell protein in 20 mM Tris-HCl (pH 8.0), with 100 mM NaCl.

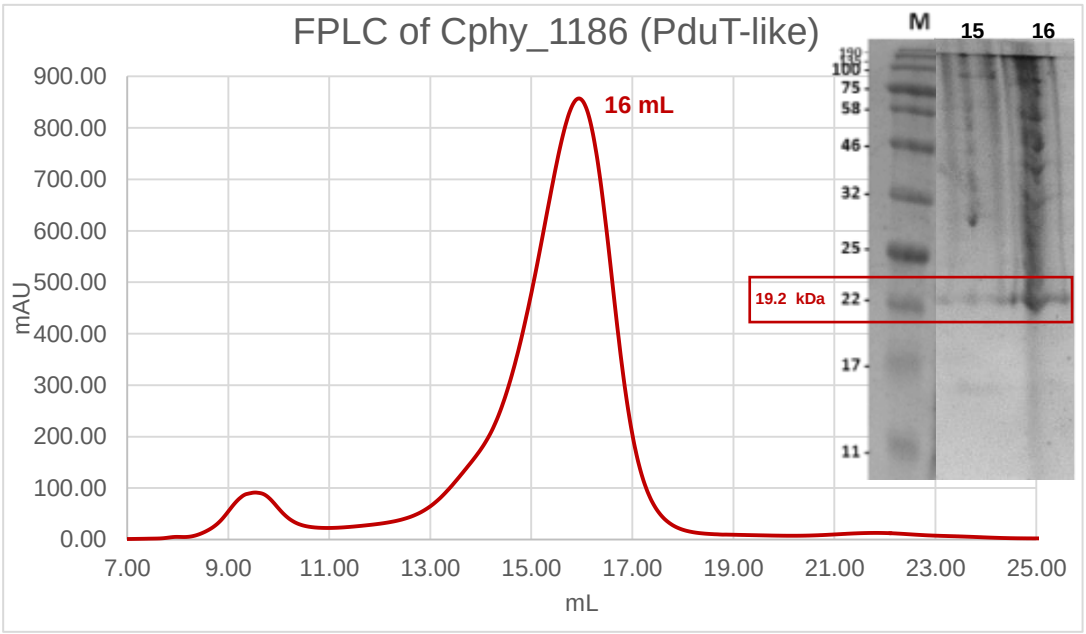


Figure 3.30: Further purification of Cphy_1186 by gel filtration chromatography. A G200 elution profile of the shell protein Cphy_1186 with a peak volume at 16 mL with the insert showing an SDS-PAGE gel of the fractions at 15 and 16 mL. The protein sizes are shaded with red colour.

The Cphy_1186 peak at 16 mL equates to a protein of approximately three times the size of Cphy_1186 **(Table 3.9)**.

Table 3.9: Peak and calibration values of Cphy_1186 shell protein. The protein size of Cphy_1186 is without the Hexa-His tag.

Proteins	Size/ kDa	Peak volme/ mL	Peak size/ kDa	Number of subunits
Cphy_1186	19.2	16	60	3

3.2.4 Engineering empty bacterial microcompartment Shell

As previously shown, the *Citrobacter freundii* bacterial microcompartments associated with 1,2-PD degradation can be successfully transferred into *E. coli* to allow the generation of recombinant BMCs. Moreover, when just the shell proteins are coordinately expressed empty BMCs are generated (Parsons *et al.*, 2010). Here, the same goal of trying to produce empty BMCs was attempted through the coordinated expression of the shell proteins from the *C. phytofermentans* fucose/rhamnose BMC operon (Cphy_1176, 1180, 1181, 1182, 1184, 1186). The genes encoding the six shell proteins were ligated into the plasmid pET3a through the **'Link and Lock' (Section 2.3.6.3)** cloning technology (McGoldrick *et al*., 2005). The resultant recombinant plasmid (pET3a-Cphy_1176, 1180, 1181, 1182, 1184, 1186) was transformed into *E. coli* BL21 Star™ (DE3)pLysS and the transformants were grown at 37°C in LB media with addition of appropriate antibiotics, ampicillin and chloramphenicol (a control of pET3a was also grown). One of these transformants was grown and incubated until the OD_{600} reading was approximately 0.6 **(Section 2.2.7)**. At this point protein overproduction was

induced by addition of IPTG prior to the culture being incubated overnight at 18°C. After overnight protein induction the culture was harvested by centrifugation and any macromolecular protein complexes that had formed was purified using yeast protein extraction reagent (Y-PER™) **(Figure 3.31) (Section 2.4.7)**. The culture was subject to the purification protocol that had been developed for recombinant empty BMCs from *E. coli*.

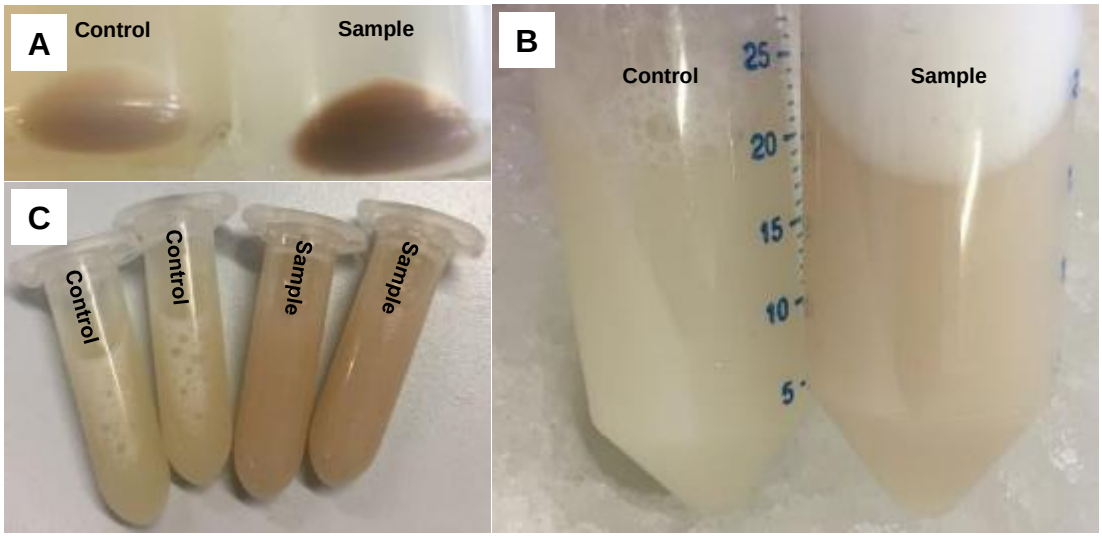


Figure 3.31: Y-PER™ purification process of the recombinant BMCs. (A) cell pellets of the control and sample after centrifugation. (B) Suspension of the control and sample pellets with 10 mL Y-PER™ + Protease inhibitor + 5 µL Benzonase Nuclease. (C) Final resuspension in 20 mM Tris-HCL pH8 + 20 mM NaCl.

Protein fractions were analysed by SDS-PAGE **(Figure 3.32)**. Two obvious bands were observed around 22 kDa in lane P_1 and P_2 (pellet 1 and pellet 2 respectively) and these correspond to the largest shell proteins of the *C. phytofermentans* fucose/ rhamnose BMC, Cphy_1186 and Cphy_1182. In addition, some lanes showed two extra wide bands between 11 - 13 kDa,

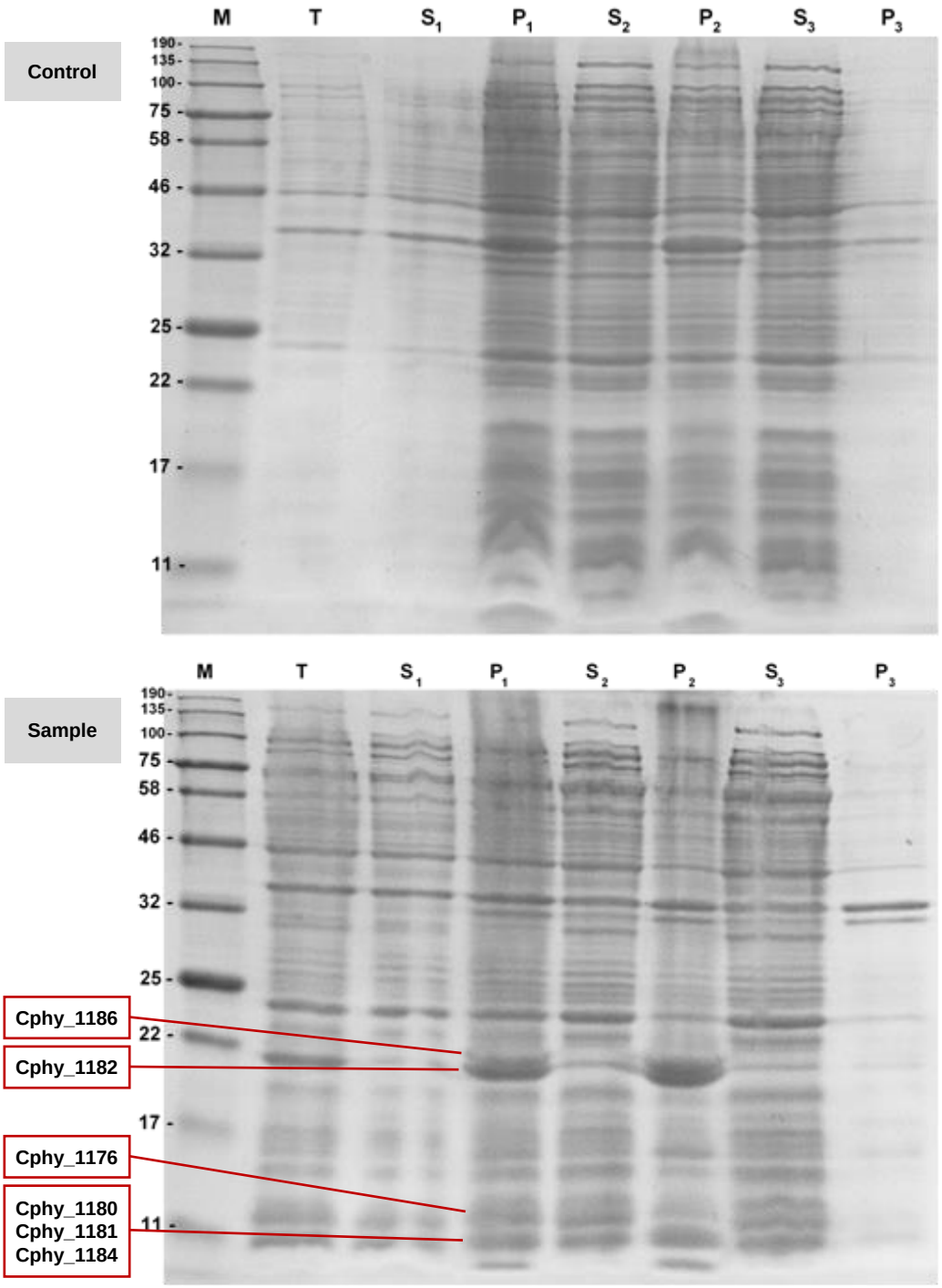


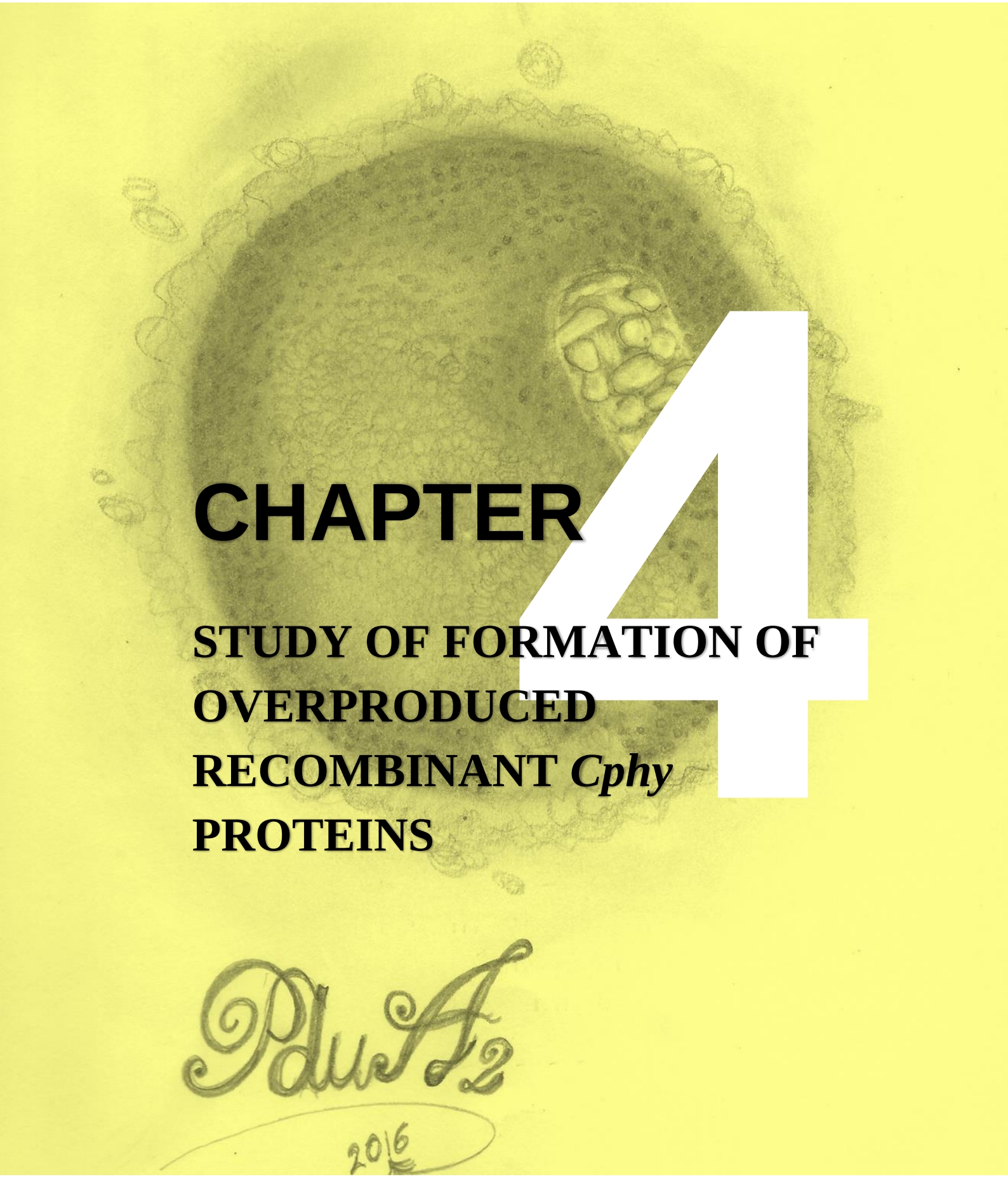
Figure 3.32: 15 % SDS-PAGE gel showing purification process. Fractions of total (T), soluble (S), insoluble (P) of overexpressed pET3a (control) and pET3a-Cphy_1176, 1180, 1181, 1182, 1184, 1186 proteins. The protein was harvested from a 250 mL culture. The expected bands are highlighted.

which likely represent of the smaller shell proteins Cphy_1176, 1180, 1181, and 1184. These bands were not present in a control SDS-PAGE of an *E.*

coli strain containing only pET3a. This result suggests that although many of the *C. phytofermentans* shell protein had been successfully produced, indicating a potential for BMC formation. The presence of BMCs needs to be confirmed by thin sectioning of bacteria and transmission electron microscopy.

3.3 Conclusion

All six shell proteins of *C. phytofermentans Cphy* were successfully overproduced recombinantly in *E. coli*. All were purified successfully through a combination of IMAC, PD-10, and FPLC chromatography, with the exception of Cphy_1176, which precipitated during concentration. For the most part the shell proteins were soluble and SDS-PAGE showed strong bands with few contaminants. The gel filtration results suggested that Cphy_1180, Cphy_1181, and Cphy_1182 formed hexamers. While, Cphy_1184, and Cphy_1186 formed pentamers or trimers respectively. In addition, cloning of all these six genes within a single plasmid led to the production of most of the proteins as determined by SDS-PAGE. Such an approach could lead to the assembly of a recombinant empty BMC.



CHAPTER 4

STUDY OF FORMATION OF OVERPRODUCED RECOMBINANT *Cphy* PROTEINS

4.1 Introduction

Although *C. phytofermentans* is known to form BMCs, no studies have been undertaken to look at the individual components of the structure, or to try and produce the BMC recombinantly. The aims of the research reported in this chapter are to express the six individual shell proteins of *C. phytofermentans* in *E. coli* in order to investigate the ability of the proteins to form filaments or sheets by transmission electron microscopy (TEM). The possibility of forming empty BMCs was also investigated by co-expression of the genes for the six shell proteins. It was hoped that such approaches would provide a greater understanding about BMC composition and help to define the minimum number of shell proteins necessary for empty microcompartment assembly.

The required genes were cloned in pET3a and pLysS plasmids and expressed into *E. coli* strain BL21*(DE3)pLysS. The recombinant cells were then characterised for the presence of internal structures such as filaments by TEM.

The study was initiated by analysing the individual production of the six shell proteins (Cphy_1176, 1180, 1181, 1182, 1184, 1186). In comparison to the Pdu shell proteins, Cphy_1180 and Cphy_1182 were assigned as PduA-like and are therefore likely to be major components of the BMC. Cphy_1176, 1181, 1184 and 1186 are likely to be more minor shell proteins as they display similarity to PduU, PduK, PduN and PduT respectively. Deletion of shell proteins such as PduU and PduT does not appear to impact BMC structure formation itself, but their absence does reduce cell growth under conditions when BMC formation is required (Cheng *et al.*, 2011). The

absence of major shell proteins impacts significantly upon organelle formation (Chun *et al.*, 2014).

The strains with recombinant plasmids (pET3a-Cphy_1176, pET3a-Cphy_1180, pET3a-Cphy_1181, pET3a-Cphy_1182, pET3a-Cphy_1184 and pET3a-Cphy_1186) were expressed in order to examine structures that formed within the cells. In addition, a strain with a recombinant plasmid encoding all six shell genes (Cphy_1176-1180-1181-1182-1184-1186) in the same order as exists in the wild-type genome was expressed under different growth conditions in order to examine its ability to form empty BMCs. Strains with different combinations of genes were studied in order to try to identify the least number of genes important for BMC assembly and to reveals some of their functional aspects. This was achieved by constructing plasmids with varying numbers of shell-protein genes, including Cphy_1176-1180-1181-1182-1184, Cphy_1176-1180-1181-1182, Cphy_1176-1180-1181, Cphy_1176-1180, Cphy_1184-1176 and Cphy_1184-1186.

4.2 Results

4.2.1 Expression of individual shell proteins

4.2.1.1 Characterisation of Cphy_1176

When produced in *E. coli*, the Cphy_1176 protein was found to be largely insoluble when cell extracts were analysed by SDS-PAGE. The small amount of soluble protein was, however, found to disappear during the concentration process presumable due to aggregation, and hence the protein

could not be used for crystallisation and structural studies. However, this shell protein has more than 50 % sequence similarity to the *S. enterica* PduU **(Figure 4.1)**. PduU forms a hexamer and is classified as a circularly permuted single-BMC-domain protein (Chun *et al.*, 2014). For TEM analysis

```
# Identity:      52/123 (42.3%)
# Similarity:    71/123 (57.7%)
# Gaps:          11/123 ( 8.9%)
# Score: 223.5


Cphy_1176          1 ----MTTEDKLRIVQELVPGKQISLAHIIAAPDPLLMEKLCFQKENERMK     46
                       .||:   |::||.|||||::|||:||.|...|.:||..|.    ..
Sen_PduU           1 MERQPTTD---RMIQEYVPGKQVTLAHLIANPGKDLFKKLGLQD----AV     43

Cphy_1176         47 AAIGILTMSPAETVIIGADLALKASGVTLQNVDYASGTLVFTGTVSEVEA     96
                     :||||||::|:|..||..|:|.|:..|.:...:|..:|.:|.||.||.||.
Sen_PduU          44 SAIGILTITPSEASIIACDIATKSGAVEIGFLDRFTGAVVLTGDVSAVEY     93

Cphy_1176         97 AMNAVVEYSNRTLSFTVCDITRT    119
                     |:..|.......:.||.|.||||
Sen_PduU          94 ALKQVTRTLGEMMQFTTCSITRT    116
```

Figure 4.1: Sequence alignment of *C. phytofermentans* Cphy_1176 and *S. enterica* PduU. This Figure was produced using EMBOSS Needle.

the Cphy_1176 shell protein was induced within the relevant strain by the addition of 200 µM IPTG and grown overnight at 18°C. Cells were harvested, embedded in resin, sectioned and stained as described in **Chapter 2**. Thin sections of the strain overproducing Cphy_1176 were visualised by TEM. The sections show that Cphy_1176 forms a number of different shapes which appear to be filamentous in nature. These structures appear to interfere with cell division **(Figure 4.2)**. The relatively round and elongated structures are between 100 - 200 nm in diameter which is comparable to BMCs.

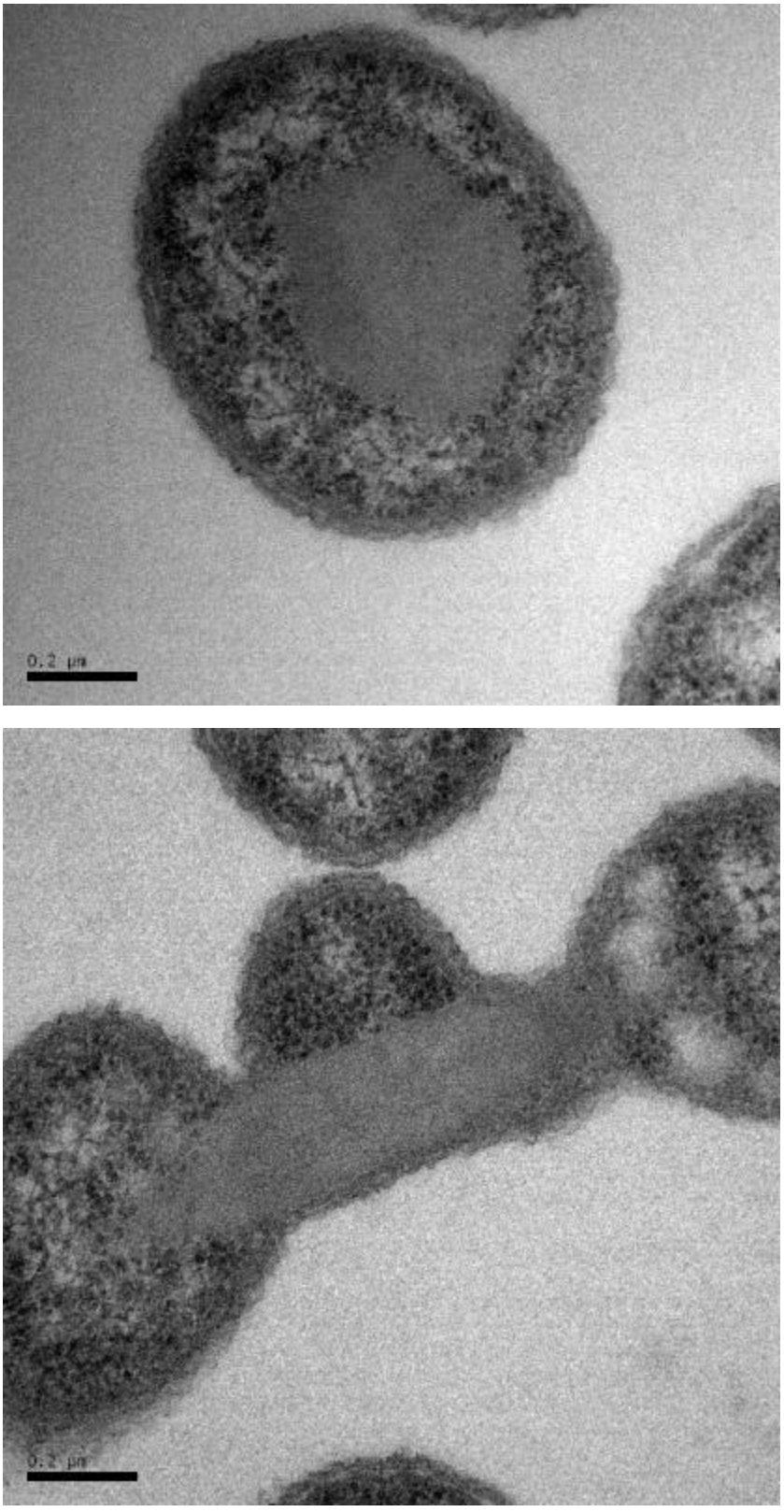


Figure 4.2: Thin section of *E. coli* overproducing Cphy_1176 (PduU-like) protein. The Image on the top shows a transverse section. On the bottom is an image of longitudinal section through the cell.

4.2.1.2 Characterisation of Cphy_1180

The Cphy_1180 shell protein has 50 % similarity to the *S. enterica* and *C. freundi* PduA protein. In addition, Cphy_1180 has 50 % similarity to the other shell protein Cphy_1182 (not including the N-terminal extension) **(Figure 4.3)**.

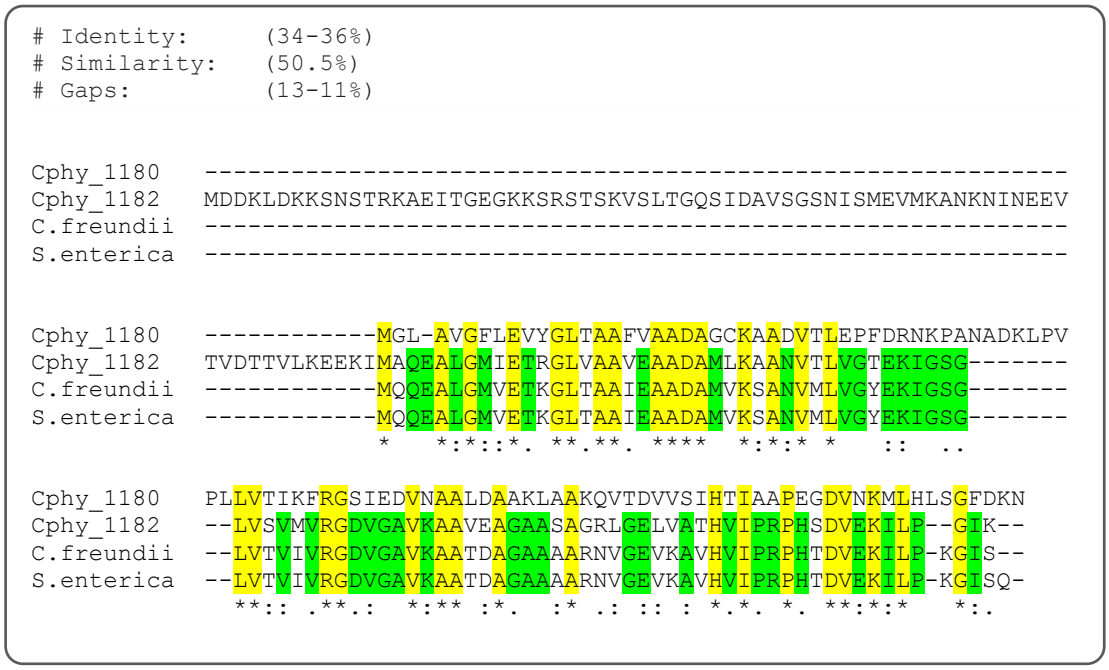


Figure 4.3: Multiple sequence alignment of Cphy_1180, Cphy_1182, *C. freundii* PduA, *S. enterica* PduA. Similarity between the sequences is shaded for easily identification. This Figure was produced using Clustal Omega (Sievers and Higgins, 2017).

When Cphy_1180 was overproduced in *E. coli*, no significant structures were observed but rather polar inclusion bodies were present in more than 40 % of the cells **(Figure 4.4)**.

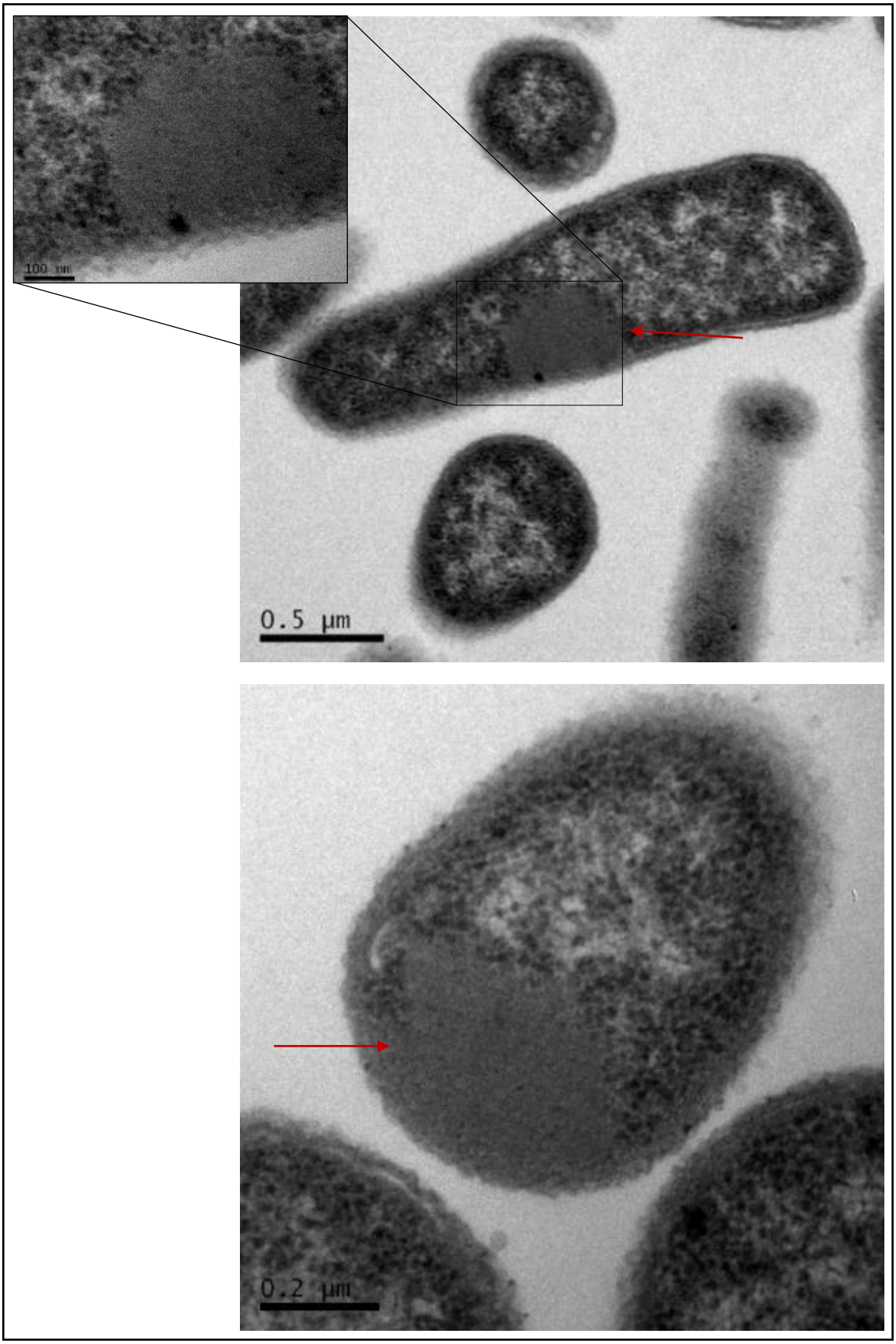


Figure 4.4: Thin section of *E. coli* overproducing Cphy_1180 (PduA-like) protein. Red arrows indicate resulting structures. The Image on the bottom shows a cell in transverse orientation. On the top is an image of longitudinal section through the cell section with magnified view.

4.2.1.3 Characterisation of Cphy_1181

Cphy_1181 has around 34 % sequence similarity to the *S. enterica* and *C. freundii* PduK. In addition, Cphy_1181 has 30 % similarity to the other shell protein Cphy_1182 of *C. phytofermentans* (including the N-terminal extension) **(Figure 4.5)**.

```
Cphy1181    ------------------------------------------------------------
C.freundii  ------------------------------------------------------------
S.enterica  M-----------------------------------------------------AN-------
Cphy1182    MDDKLDKKSNSTRKAEITGEGKKSRSTSKVSLTGQSIDAVSGSNISMEVMKANKNINEEV


Cphy1181    -------------MGKSLGFIEISGVTAAIDALDIMCKTAGVELVTWERKMGGRLVTIIV
C.freundii  -------------MKQSLGLLEVSGLALAISCADVMAKAASITLVGLEKTNGSGWTVIKI
S.enterica  -------KEHR--VKQSLGLLEVCGLALAISCADIMAKSASITLLALEKTNGSGWMVIKI
Cphy1182    TVDTTVLKEEKIMAQEALGMIETRGLVAAVEAADAMLKAANVTLVGTEKI-GSGLVSVMV
                         ::**::*  *:. *:.. * * *:*.: *:  *:  *.    : :


Cphy1181    SGSVSEVTQAVESAVNQAIKKP--VAHAVIANPHEEVLRLVDISASRMKPQN--------
C.freundii  IGDVASVQAAISTGANLAELRNGLVAHRVISRPGEGILSHSVPVQPEPAPEPIPYLSQEQ
S.enterica  AGDVASVQAAITTGAQFAEQRNGLVAHKVIARPGEGILLAETPSPSVIEPEP---EASEI
Cphy1182    RGDVGAVKAAVEAGAASAGRLGELVATHVIPRPHSDVEKIL--------------------
             *.*. *  *: :..  *      **  **..* . :


Cphy1181    EEM-----------------------------------------------
C.freundii  EEMATEIALPEAAQQAELISCNLCLDPACPRQKGEPRTLCLHSGKRGEA
S.enterica  ADVVSEAPAEEAPQESELVSCNLCLDPKCPRQKGEPRTLCIHSGKRGEA
Cphy1182    -------------------------------PGIK----------------
```

Figure 4.5: Multiple sequence alignment of Cphy_1181 with Cphy_1182 and PduK proteins from different organisms. Similarity between the sequences is shaded for easily identification. This Figure was produced using Clustal Omega (Sievers and Higgins, 2017).

An *E. coli* strain containing Cphy_1181 produced unusual structures **(Figure 4.6)**. Elongated laminar sheets were observed. This is consistent with the presence of “rolled up” protein sheets, which on cross section showed a

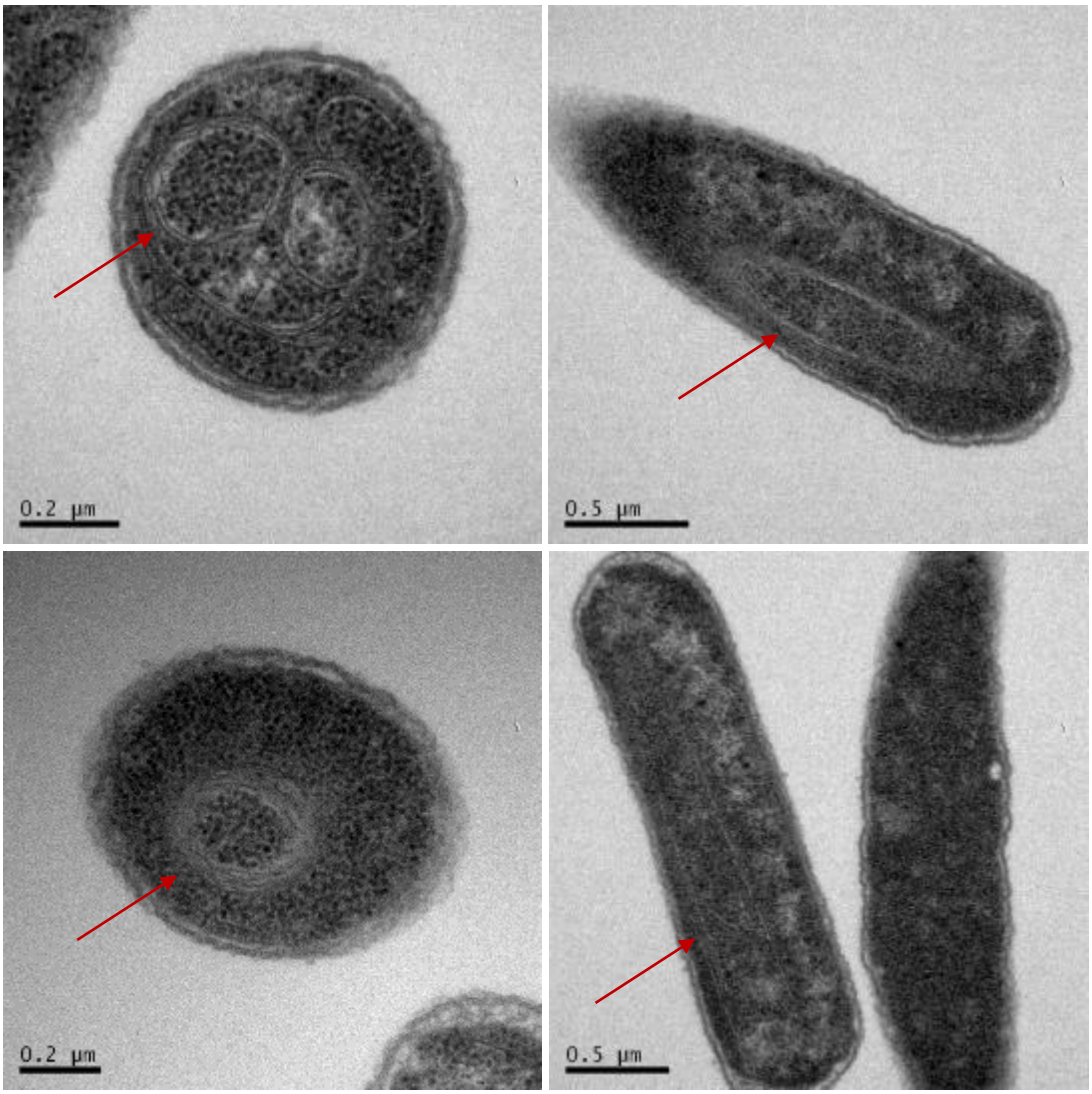


Figure 4.6: Transmission electron microscopy of native Cphy_1181 (PduK-like) protein produced in *E. coli*. Red arrows indicate resulting structures. Images on the left show transverse sections and images on the right show longitudinal sections through cells.

spiral appearance. It seems as though these elongated sheets fold to give a Swiss roll type of structure.

4.2.1.4 Characterisation of Cphy_1182

As previously mentioned, Cphy_1182 protein has similarity to PduA, K, Cphy_1180 and Cphy_1181 **(Figure 4.5)**. An *E. coli* strain expressing Cphy_1182 produced higher order structures that resemble a network of

complete round rings in transverse sections, which do not display a honeycomb appearance (hexagonal lattice) as is observed with PduA expression. Cylinder (tube-like) structures can be seen in longitudinal sections **(Figure 4.7)** which is consistent with filament formation. Measurement of the structures formed with Cphy_1182 indicates that the tubes are approximately 25 nm in diameter. Furthermore, the tubes appear as pairs of sectioned lines or as stacked elongated capsules, but not as PduA pairs of long lines (Pang *et al.*, 2014). It is proposed that in the cytoplasm, Cphy_1182 molecules tend to form hexamers, which in some way assemble to form nanotubes, and that differs from what seen in the crystal lattice for the same protein **(See chapter 5)**.

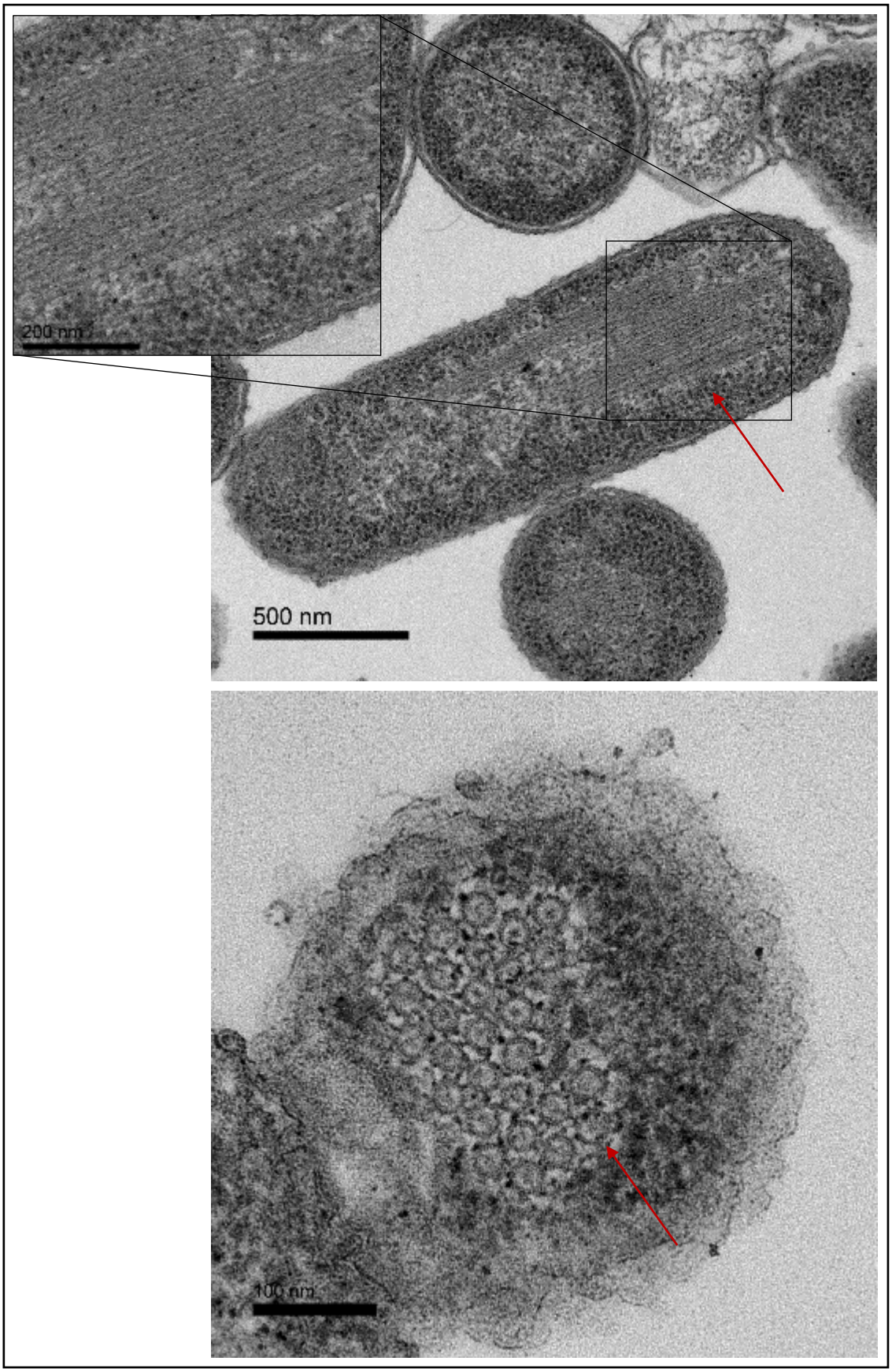


Figure 4.7: Transmission electron microscopy of *E. coli* cells expressing native Cphy_1182 (PduA-like). Red arrows indicate resulting structures. The Image on the bottom shows transverse section. On the top is an image of longitudinal section through a cell with magnified view.

4.2.1.5 Characterisation of Cphy_1184

The Cphy_1184 shell protein shares around 60 % and 54 % similarity to the *S. enterica* PduN and EutN respectively **(Figure 4.8),** the pentameric BMC

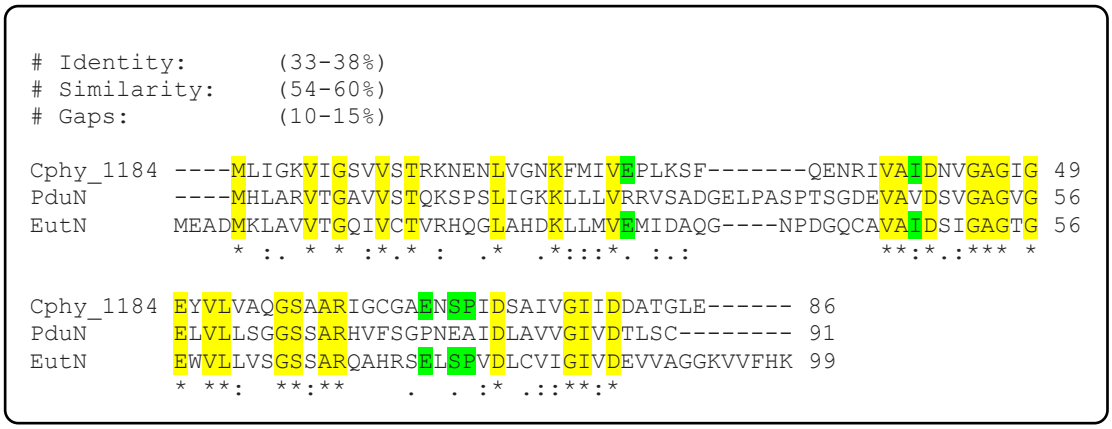


Figure 4.8: Multiple sequence alignment of Cphy_1184 with PduN and EutN proteins from *Salmonella*. Similarity between the sequences is shaded for easily identification. This Figure was produced using Clustal Omega (Sievers and Higgins, 2017).

components. This protein was recombinantly overproduced in *E. coli*, and after growth the cells were embedded, thin sectioned and analysed by TEM in the same way as for previous shell proteins. However, when the thin sections of cells overproducing Cphy_1184 were viewed, no structures were observed in any of the *E. coli* cells **(Figure 4.9)**.

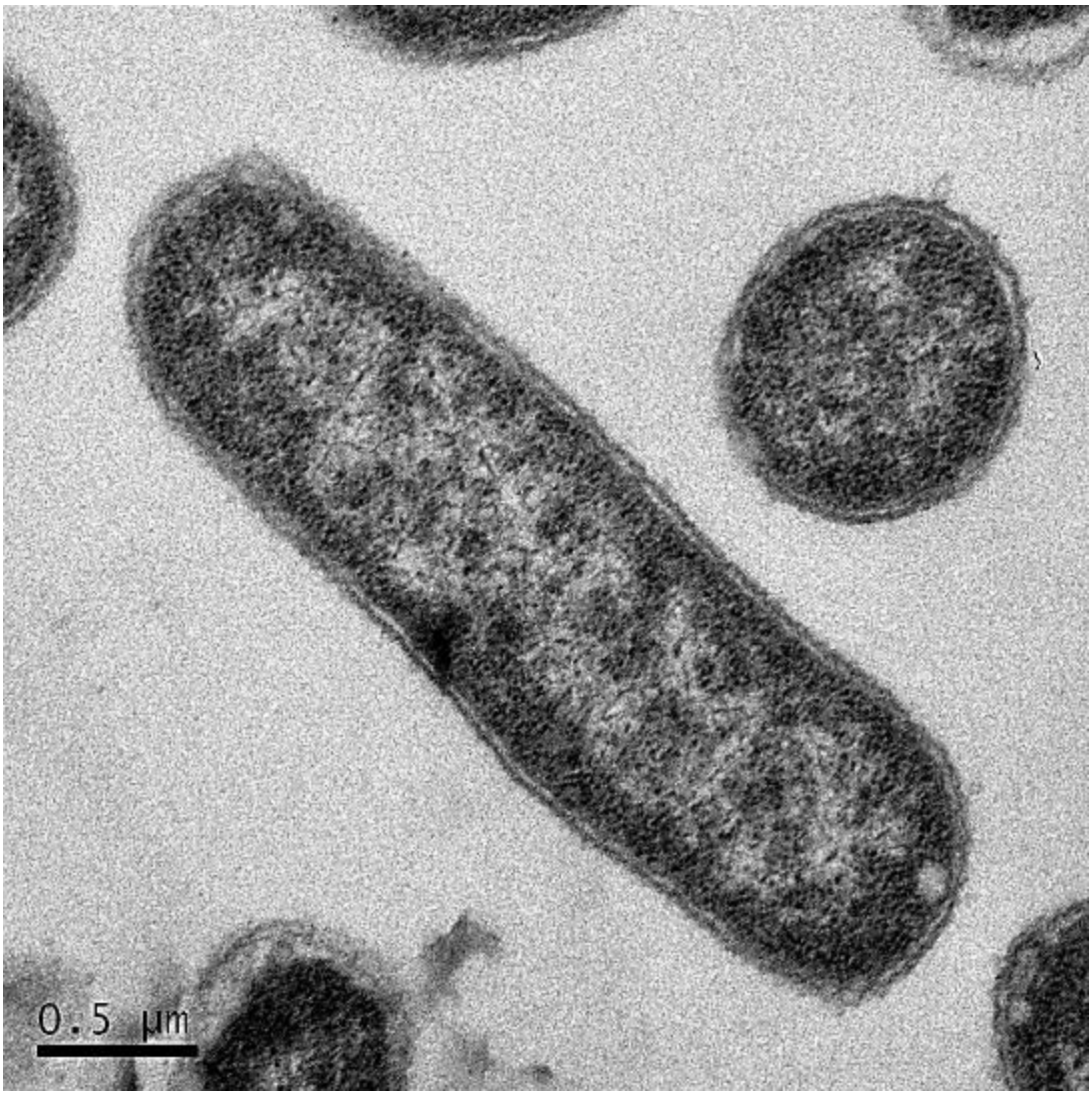


Figure 4.9: Transmission electron micrograph of *E. coli* cells expressing native Cphy_1184 (PduN-like). The cells were induced overnight with 400 µM IPTG. The cells contained no obvious structures.

4.2.1.6 Characterisation of Cphy_1186

Cphy_1186 is 66 % similar to *S. enterica* PduT **(Figure 4.10)**, the Fe-S containing shell protein. An *E. coli* strain overproducing the metalloprotein Cphy_1186 formed rolled sheet-like structures that could be observed in longitudinal sections within the cell cytoplasm, but appeared more like

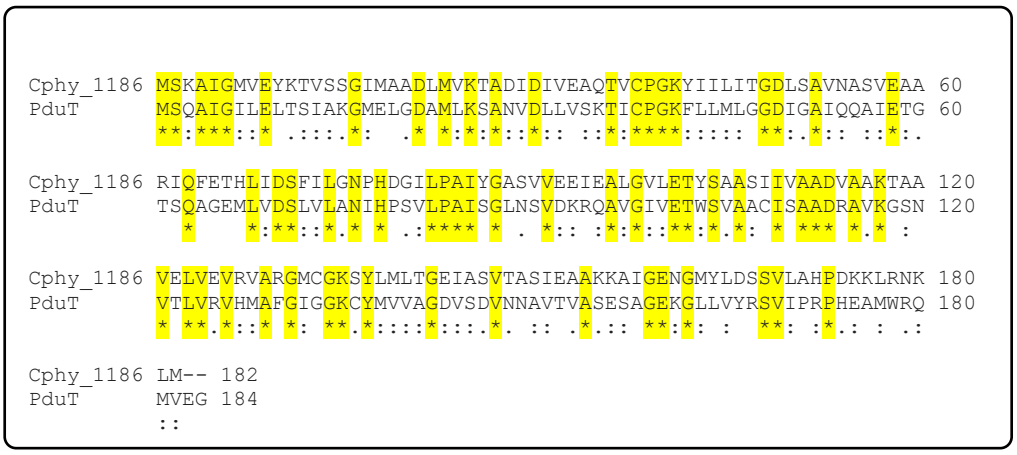


Figure 4.10: Sequence alignment of Cphy_1186 with PduT shell protein from *Salmonella*. Similarity between the sequences is shaded for easily identification. This Figure was produced using Clustal Omega (Sievers and Higgins, 2017).

circular aggregates in transverse sections **(Figure 4.11)**. These structures could be some kind of protein sheet of double stacking trimers, where the concave side facing out as seen in the crystal structure (Section 5.2.3).

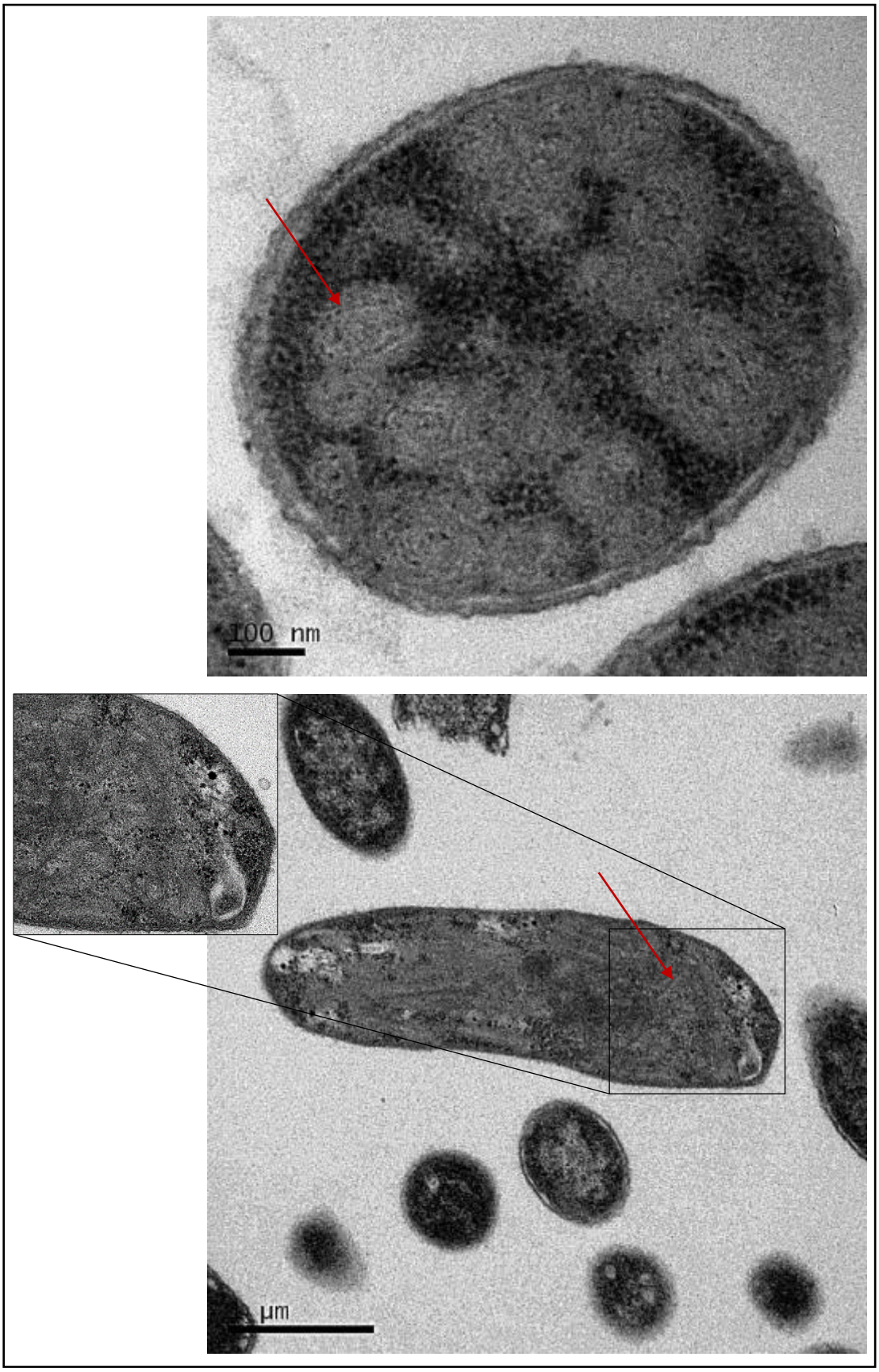


Figure 4.11: Transmission electron microscopy of *E. coli* cells expressing native Cphy_1186 (PduT-like). Red arrows indicate resulting structures. The image on the top shows transverse section. On the bottom is an image of longitudinal section through the cell with a magnified view.

4.2.2 Expression of engineered empty BMCs

The *Cphy* operon encodes six shell proteins (Cphy_1176-1180-1181-1182-1184-1186), which collectively form the outer structure of BMC shell and displays similarity to the BMC shell proteins of the Pdu microcompartment. In an attempt to form empty microcompartments, attempts were made to clone just the genes encoding the predicted shell proteins in the same order as exists in the wild-type *Cphy* operon in the expression vector to generate a synthetic shell-operon. The construct was successfully assembled and the resulting vector was transformed into *E. coli* strain BL21*(DE3)pLysS. The transformed cells were grown, induced with different concentrations of IPTG (200 µM, 400 µM) for 2 hours and were then analysed.

The *E. coli* cells were embedded, thin sectioned, negatively stained and imaged by TEM. The micrographs indicated the presence of large irregular internal cytoplasmic structures and bodies. The large semi-BMC structures that appear as protein sheets that have twisted round themselves to form a hexameric shape in a unique configuration within the cell cytoplasm with a size of about 200 nm in diameter (the structures were found in cells that had been induced with 400 µM IPTG/ 2 h) **(Figure 4.12 AB)**. In some cells, remarkable polar hexameric aggregations of varying sizes, both large and small, could be observed in cells that had been induced with 200 µM IPTG overnight **(Figure 4.12 CD)**. Some of the hexameric aggregations appear to affect the shape of the cell, as shown in **(Figure 4.12 C)**. The sizes of all these different structures in many cases exceed the size of the wild type the

wild-type *C. phytofermentans* BMC, which has an average diameter of around 100 nm (Petit *et al.*, 2013).

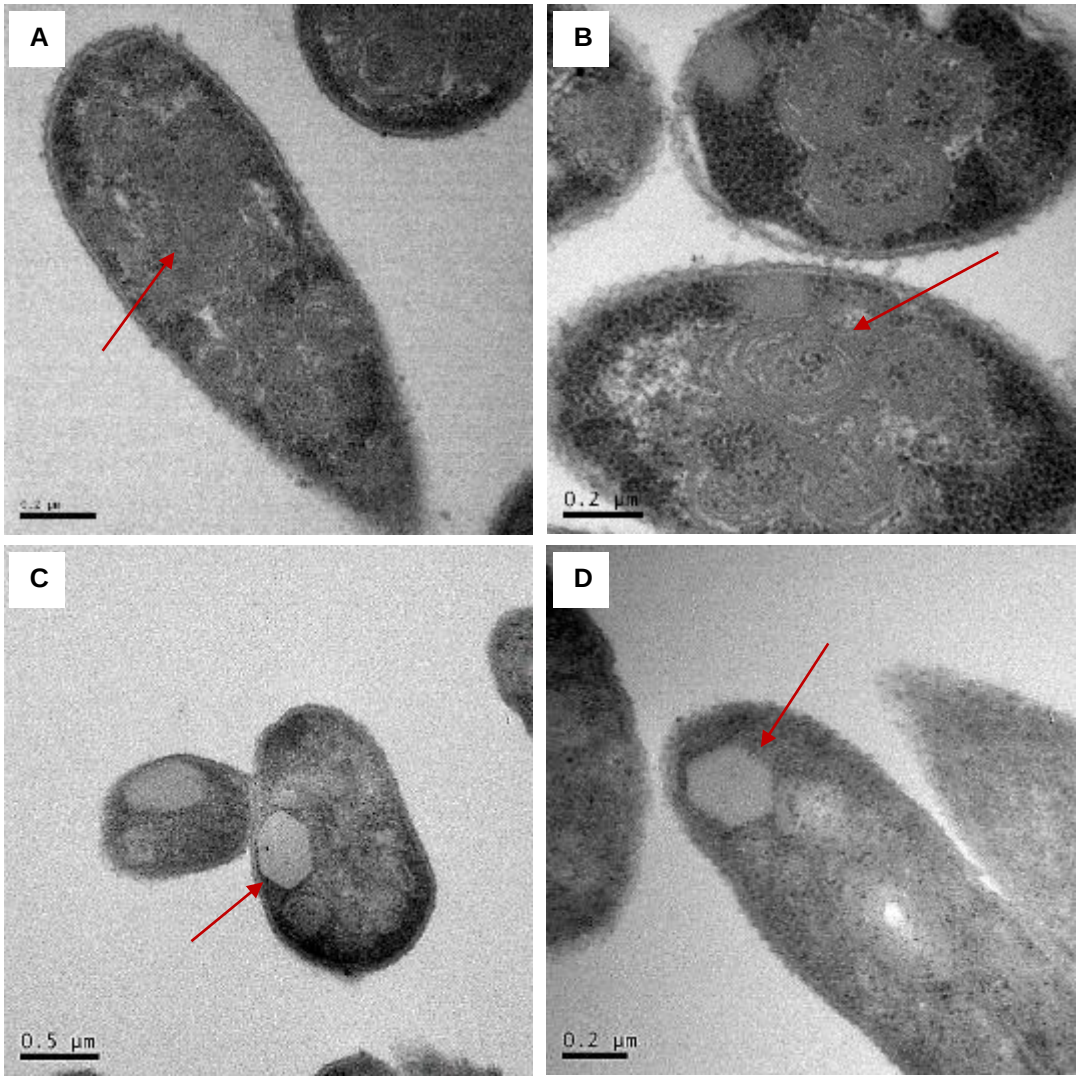


Figure 4.12: Transmission electron micrographs of *E. coli* cells expressing the engineered constructs. Red arrows indicate structures of interest. (A, B) Thin section of cell expressing empty BMCs in condition of 200 and 400 µM IPTG/ 2 h. (C, D) Thin section of cell expressing empty BMCs in condition of 200 µM IPTG overnight.

4.2.3 Purification of semi-BMCs structures

In order to characterise the structures observed in the cells expressing the six shell-protein genes, attempts were made to isolate the structures by extracting them. Extraction was achieved by using the lysis reagent (Y-PER™) as previously described **(Section 2.4.7** and **Section 3.2.4)**. The purification and isolation of the protein components of the structures was followed via SDS-PAGE **(Section 3.2.4)**. Some of the purified material was resin embedded, thin sectioned and then stained for TEM analysis. No obvious BMC structures were observed however a large amount of lipid was observed **(Figure 4.13)**. However, other purification fractions were placed on carbon coated grid and then stained before imaging by TEM, which showed some structures approximately 200 nm in diameter. These structures have some straight edges but are not strictly hexagonal **(Figure 4.14)**.

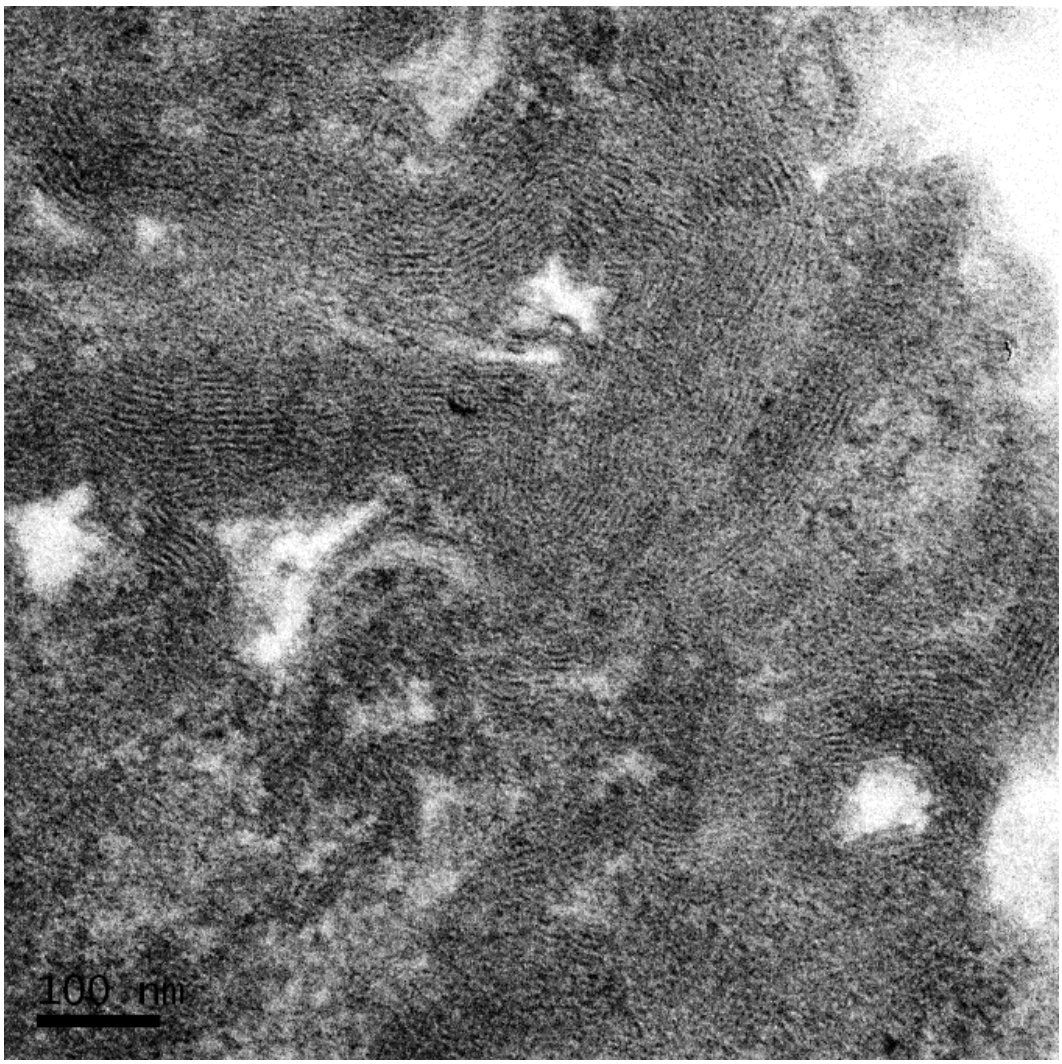


Figure 4.13: Thin sectioned and stained samples following purification. *E. coli* cells were induced overnight with 200 µM IPTG and protein was purified by (Y-PER™) method. Image indicates lipid structures.

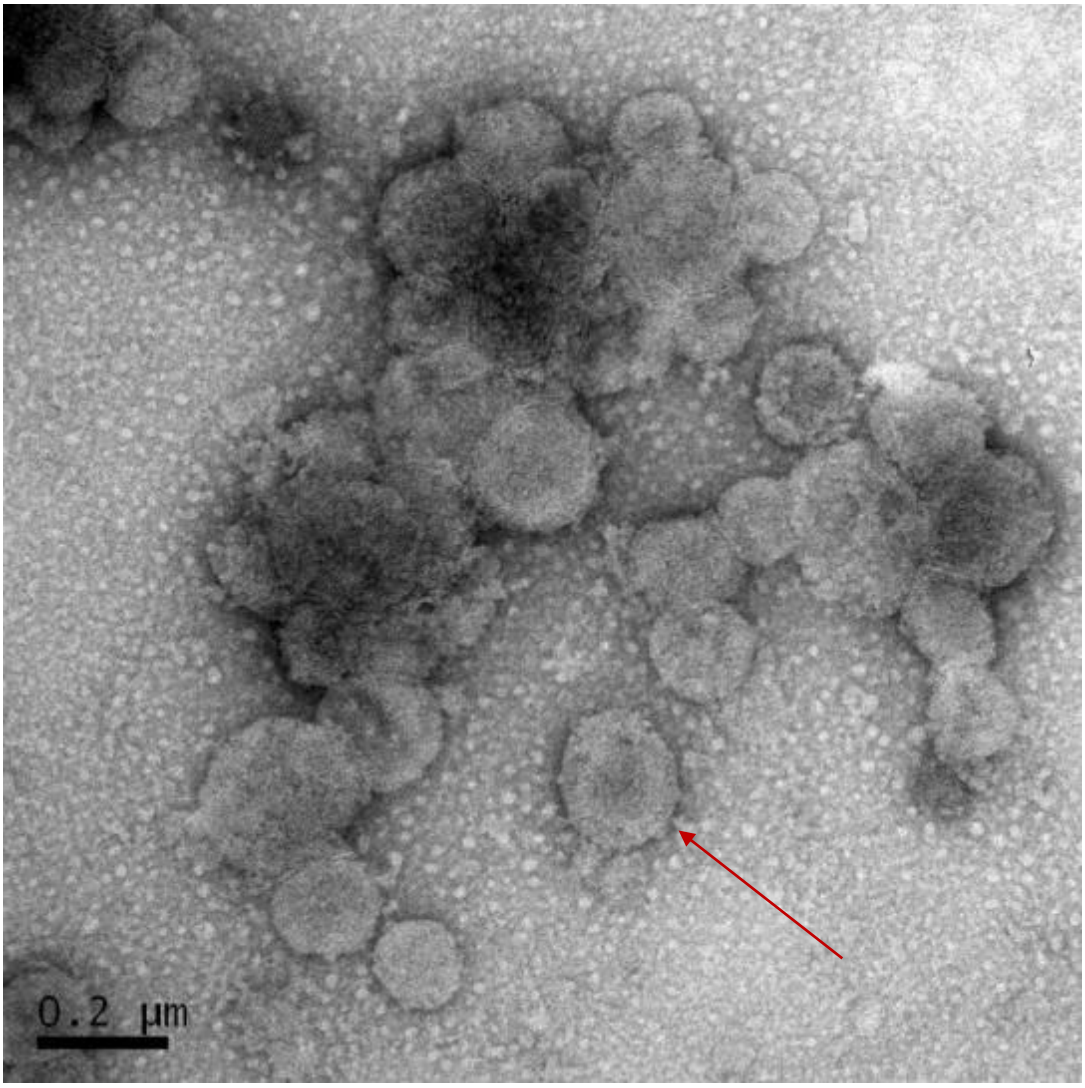


Figure 4.14: Negative stain TEM of purified empty semi-BMCs structures the red arrow indicates one of the structures. *E. coli* cells were induced overnight by 100 µM IPTG and protein was purified by (Y-PER™) method.

4.2.4 Effect of different combinations of shell proteins on organelle formation

Constructs were made with varying numbers of genes encoding for the six shell proteins in order to see what effect the various combinations had on structure formation with the cell. Constructs containing Cphy_1176-1180, Cphy_1176-1180-1181, Cphy_1176-1180-1181-1182, Cphy_1176-1180-

1181-1182-1184, Cphy_1184-1176, and Cphy_1184-1186, were constructed and transformed into *E. coli*. The resulting transformants were analysed after cells were harvested, embedded, thin sectioned and viewed by TEM. In contrast to the strain that expressed all six shell-protein genes, none of the other variants appeared to produce large microcompartment like assemblies.

4.2.4.1 Expression of Cphy_1176-1180

A cell harboring Cphy_1176-1180, when analysed by TEM after thin sectioning, revealed large homogenous structures that can be seen in almost every cell **(Figure 4.15)**. The expression also showed different geometric shapes with diameter that varied between 200 nm and 500 nm. These structures appear as some kind of ordered aggregation. They are not polar deposits but appear as a structure where protein appears to be deposited in an ordered fashion.

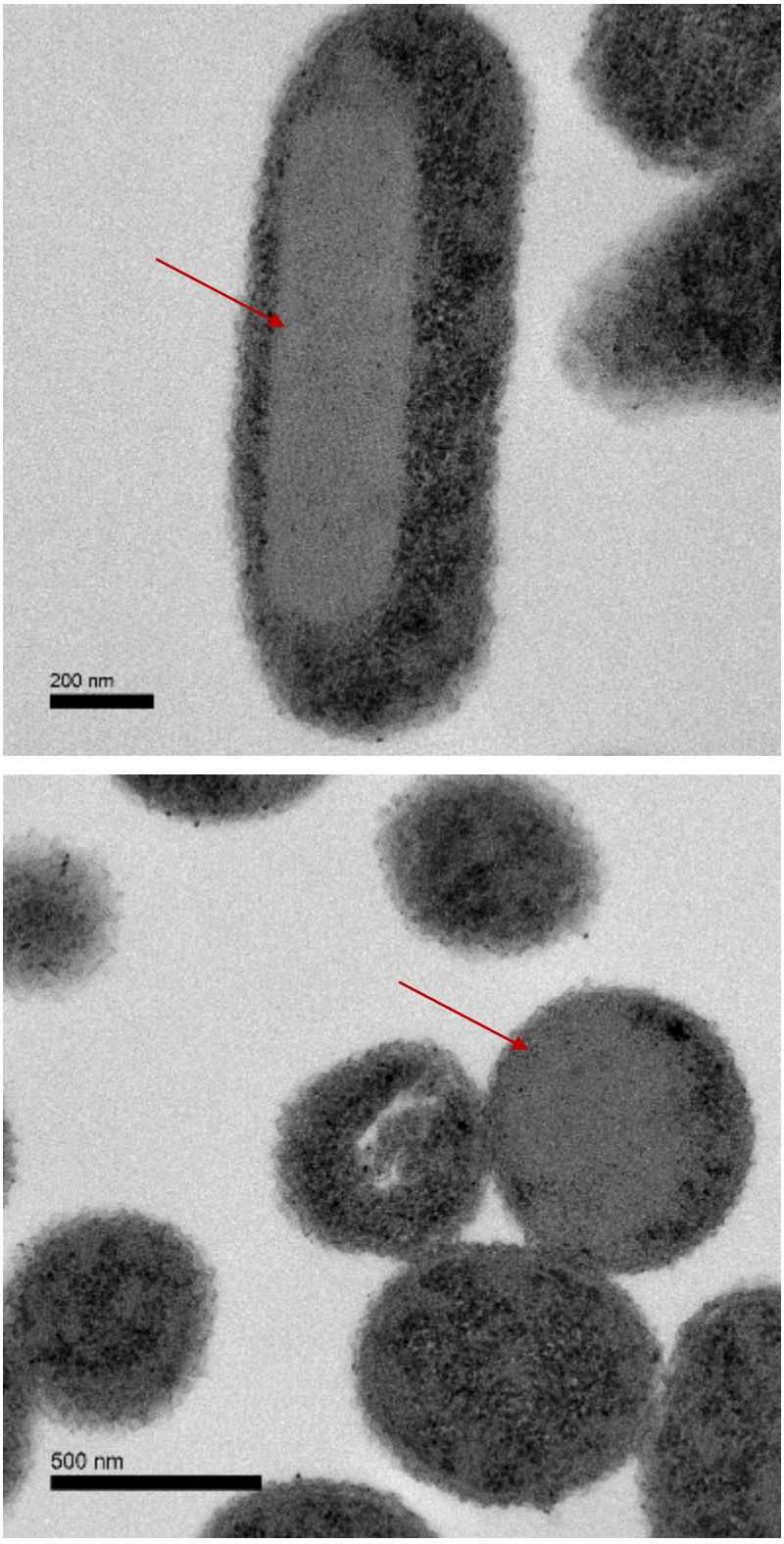


Figure 4.15: TEM analysis of *E. coli* cells expressing Cphy_1176-1180 (Ligated into pET3a) with condition of 200 µM IPTG overnight. Red arrows indicate resulting structures. On the top is a longitudinal section showing ordered aggregation. On the bottom is a transvers section showing aggregation.

4.2.4.2 Expression of Cphy_1176-1180-1181

The construct encoding three different shell proteins, Cphy_1176-1180-1181, also failed to generate microcompartments when expressed in an appropriate *E. coli* strain. Instead a single polar large geometric structure (most of the structures were hexamer in the cross-section) was formed in most of the cells **(Figure 4.16)**. It is thought that these shaped structures could represent protein crystals within the cell.

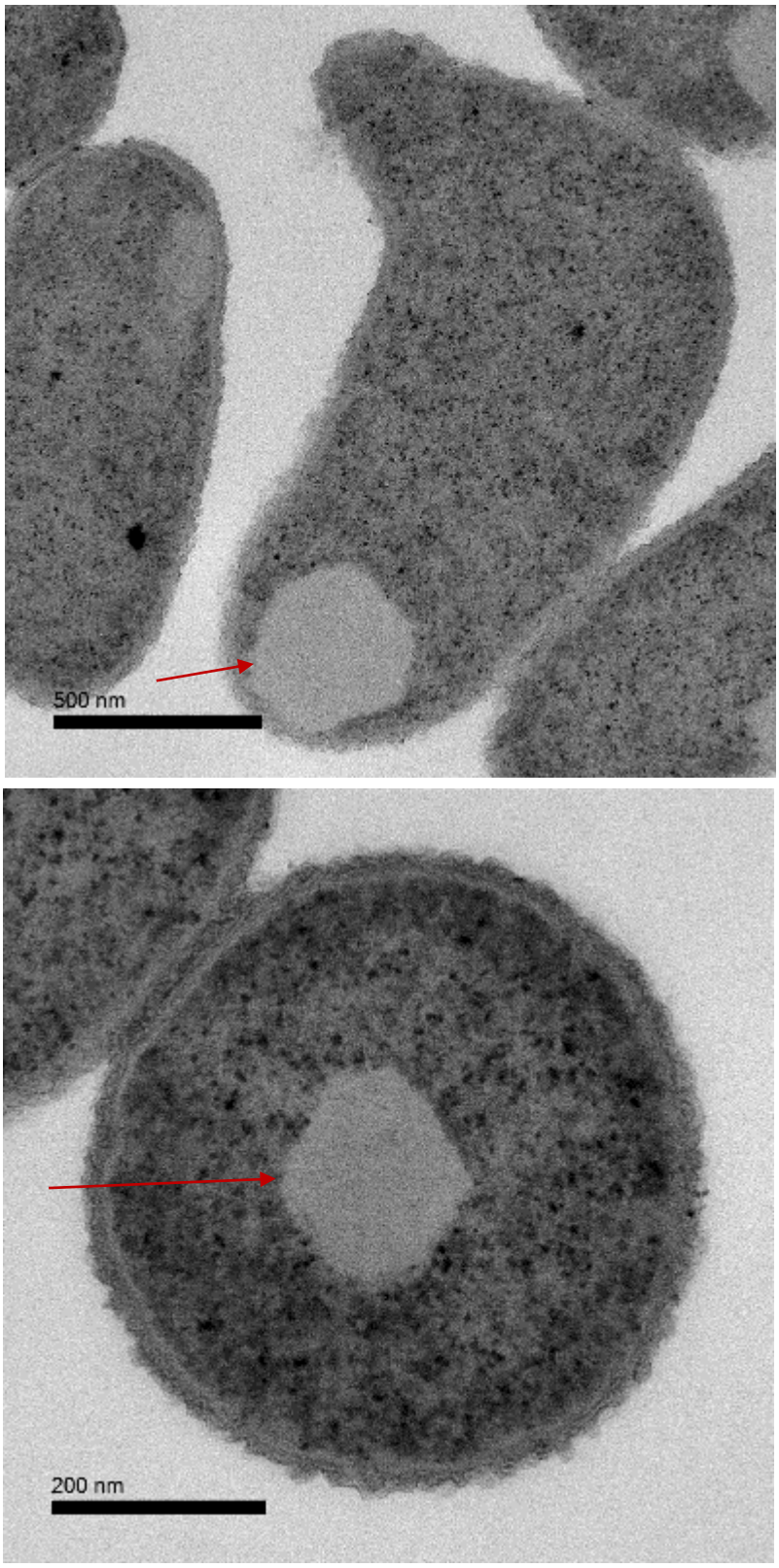
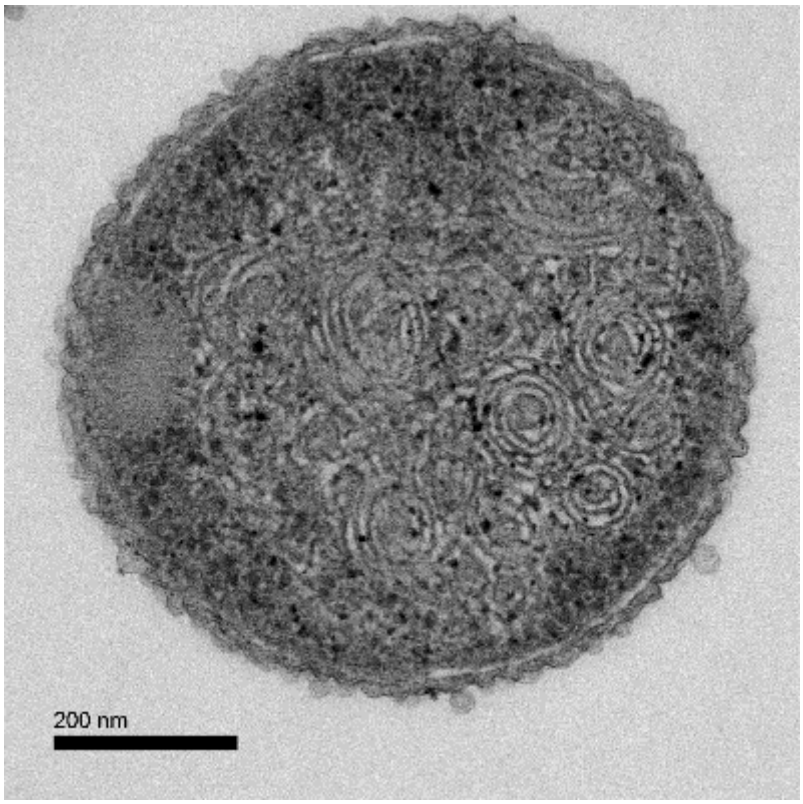


Figure 4.16: Transmission electron micrographs of *E. coli* cells expressing Cphy_1176-1180-1181 (Ligated into pET3a) with condition of 200 µM IPTG overnight. Red arrows indicate resulting structures. The image on the bottom shows transverse section, and the image on the top shows longitudinal section through a cell.

4.2.4.3 Expression of Cphy_1176-1180-1181-1182

The genes of 4 different shell proteins, Cphy_1176-1180-1181-1182, were cloned collectively into both pET3a and pLysS. The resulting plasmids were transformed into *E. coli* and the cells prepared for analysis after thin sectioning. Inclusion bodies, small twisted sheets, and nanotubes were all observed in cells that had been transformed in the pET3a vector **(Figure 4.17)**. Different structures were observed in the cells when pLysS was used as the host vector **(Figure 4.18)**. Some of these structures display hexameric shapes that sequestered some of the cytoplasm and have a size of 200 nm in diameter **(Figure 4.18D)**.



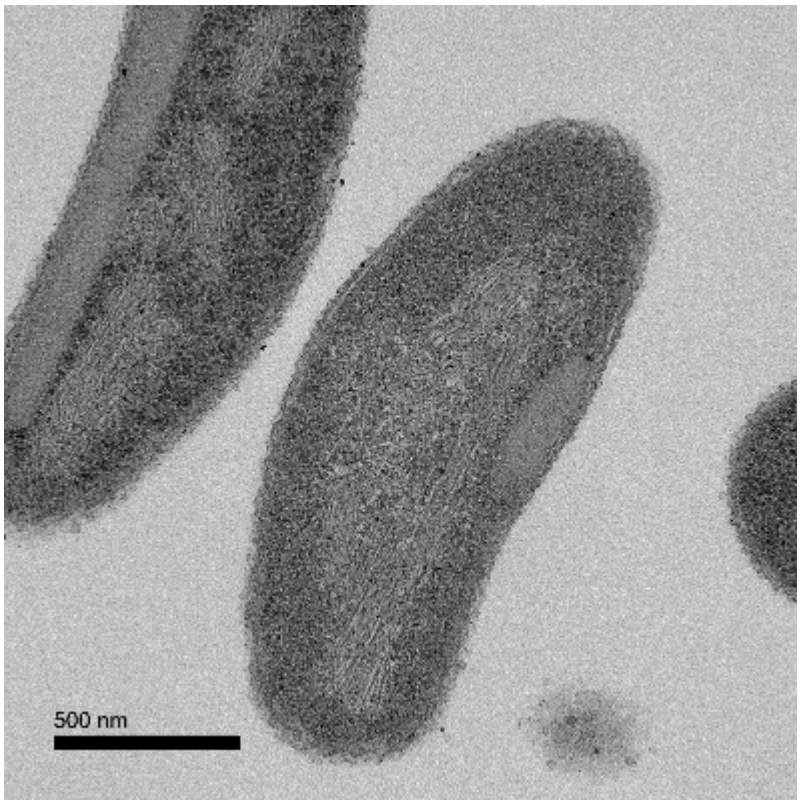


Figure 4.17: Transmission electron micrographs of *E. coli* expressing Cphy_1176-1180-1181-1182 (Ligated into pET3a) with condition of 200 µM IPTG overnight. The Image on the top shows transverse section, and the image on the bottom shows longitudinal section through a cell

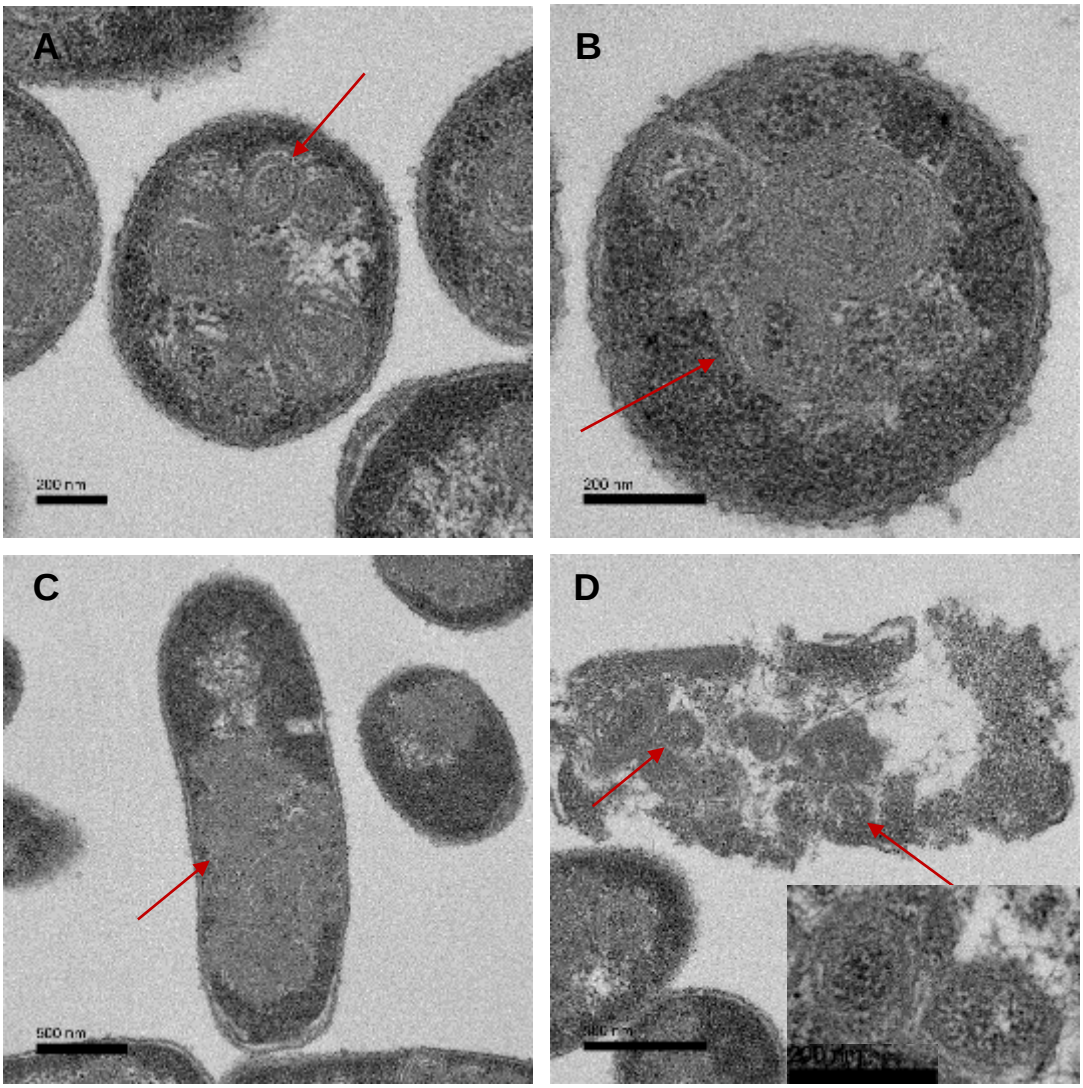


Figure 4.18: Transmission electron micrographs of *E. coli* cells expressing Cphy_1176-1180-1181-1182 (Ligated into pLysS) with condition of 200 µM IPTG overnight. Red arrows indicate resulting structures. A and B are Images that show transverse sections through the cell. C and D are images that show longitudinal sections through a cell (C) and of a lysing cell (D) with magnified view.

4.2.4.4 Expression of Cphy_1176-1180-1181-1182-1184

When cells expressing the 5-gene construct encoding Cphy_1184-1180-1181-1182-1184 were imaged a significant number of BMC-like structures between 100 - 150 nm in diameter were observed. These are of a similar size to wild-type 100 nm BMCs observed in *C. phytofermentans* and similar to purified Pdu microcompartments (Yeates *et al.*, 2010; Petit *et al.*, 2013) **(Figure 4.19A)**. Some cells produced further unusual structures that appear very long and fibrous **(Figure 4.19B)**.

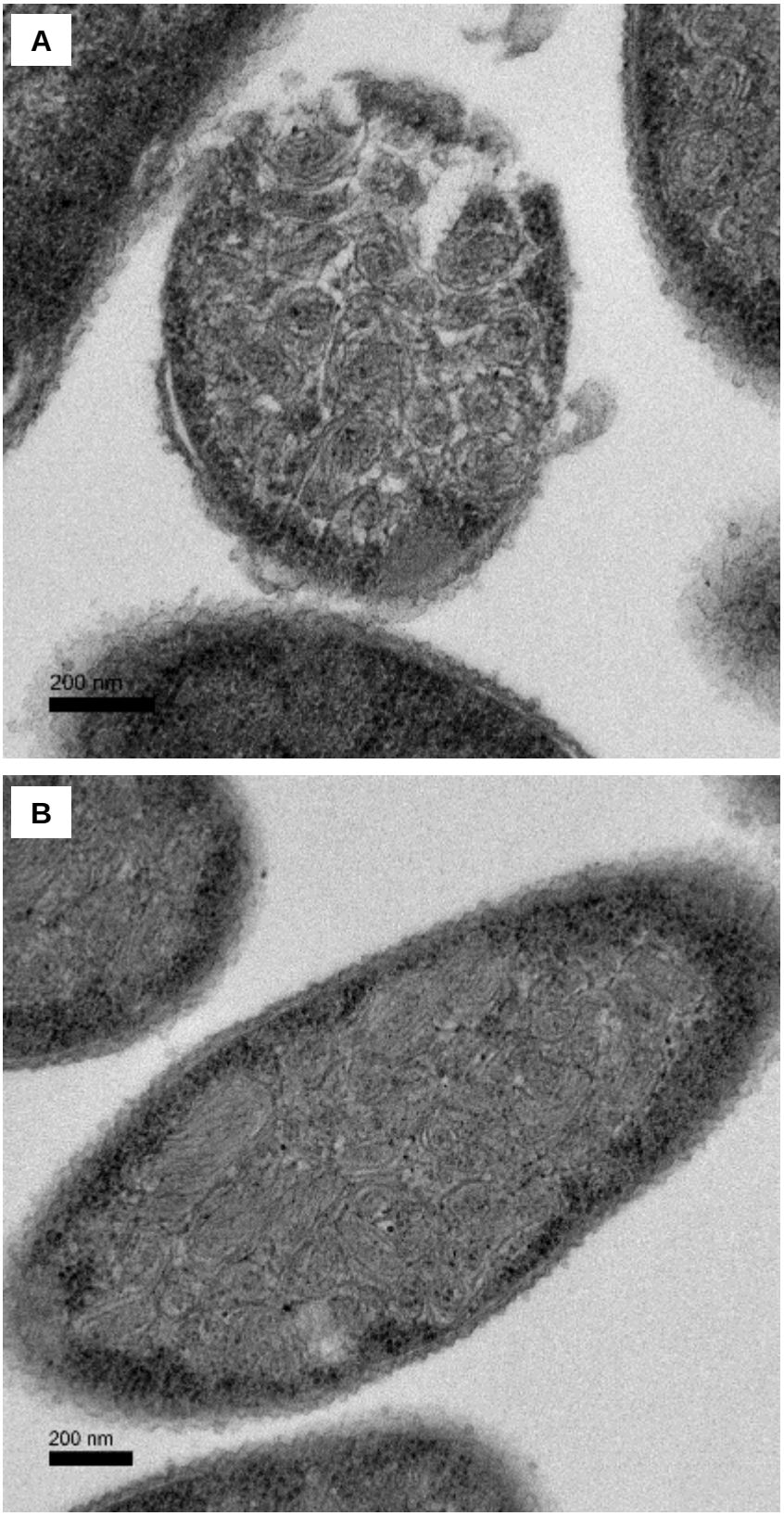


Figure 4.19: Effects of combination of Cphy_1176-1180-1181-1182-1184 (Ligated into pET3a) for the *Cphy* microcompartments formation. (A) Thin section of *E. coli* cell showing BMC-like structures from lysed cell. (B) Thin section of *E. coli* cell showing elongated structures. Cells were grown with condition of 200 µM IPTG overnight.

4.2.4.5 Expression of Cphy1184-1176

The co-expression of two shell-proteins, Cphy_1184-1176 resulted in similar large geometric (hexamer) cytoplasmic aggregates to those seen in **Section 4.2.4.2** were observed. This strain also failed to form intact normal BMCs structures **(Figure 4.20)**. The structures have a very even electron density

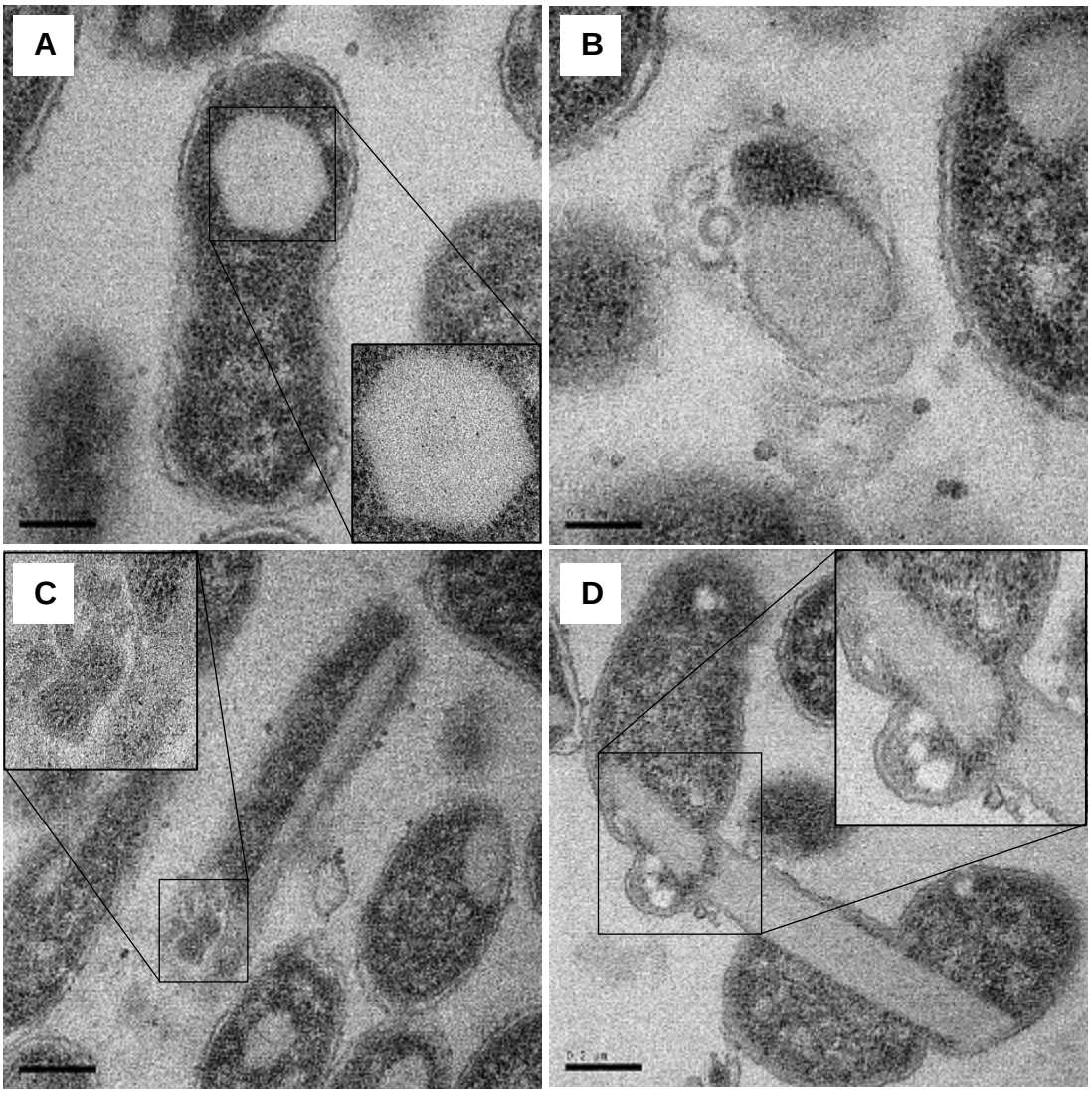


Figure 4.20: TEM analysis of *E. coli* cells expressing Cphy_1184-1176 (Ligated into pET3a) for the *Cphy* microcompartment. (A) Thin sections of *E. coli* cell expressing the delimiting polar body. (B) Thin sections of *E. coli* lysed cell expressing the polar body. (C) Thin section of *E. coli* cells forming BMC-like structure. (D) Thin section of *E. coli* cells expressing elongated shaped aggregation. Cells were induced with 200 µM IPTG overnight.

suggesting that they may represent internal cellular protein crystals. The very long structure observed in **Figure 4.20 D** appears to be fractured, consistent with this structure being crystalline. The hexamers have a diameter between 300 nm and 500 nm.

4.2.4.6 Expression of Cphy1184-1186

When Cphy_1184-1186 were overproduced in *E. coli*, longitudinal and transverse sections showed the presence of aggregated material, some, but not all, of which appeared to be as polar deposits **(Figure 4.21)**. In contrast to the longitudinal sections of *E. coli* overproducing Cphy_1186 alone **(Figure 4.11)** there is no evidence of filament-like structures. It would appear that Cphy_1184 and Cphy_1186 interact to generate a non-specific aggregate.

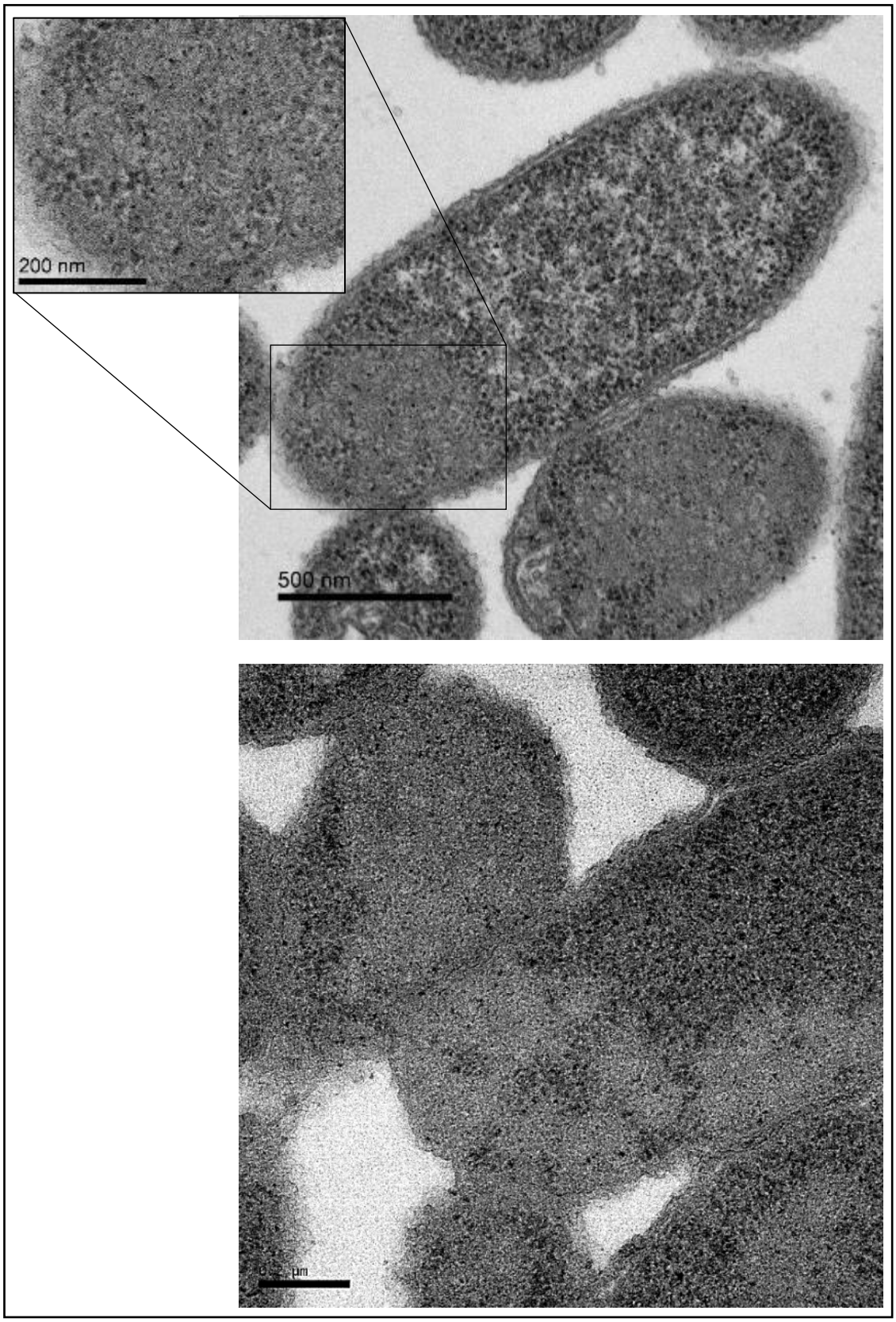


Figure 4.21: TEM analysis of *E. coli* cells expressing Cphy_1184-1186 (Ligated into pET3a) for the *Cphy* microcompartments. Cells were grown with condition of 200 µM IPTG overnight. Similar structures were observed in over 60 % of the cells.

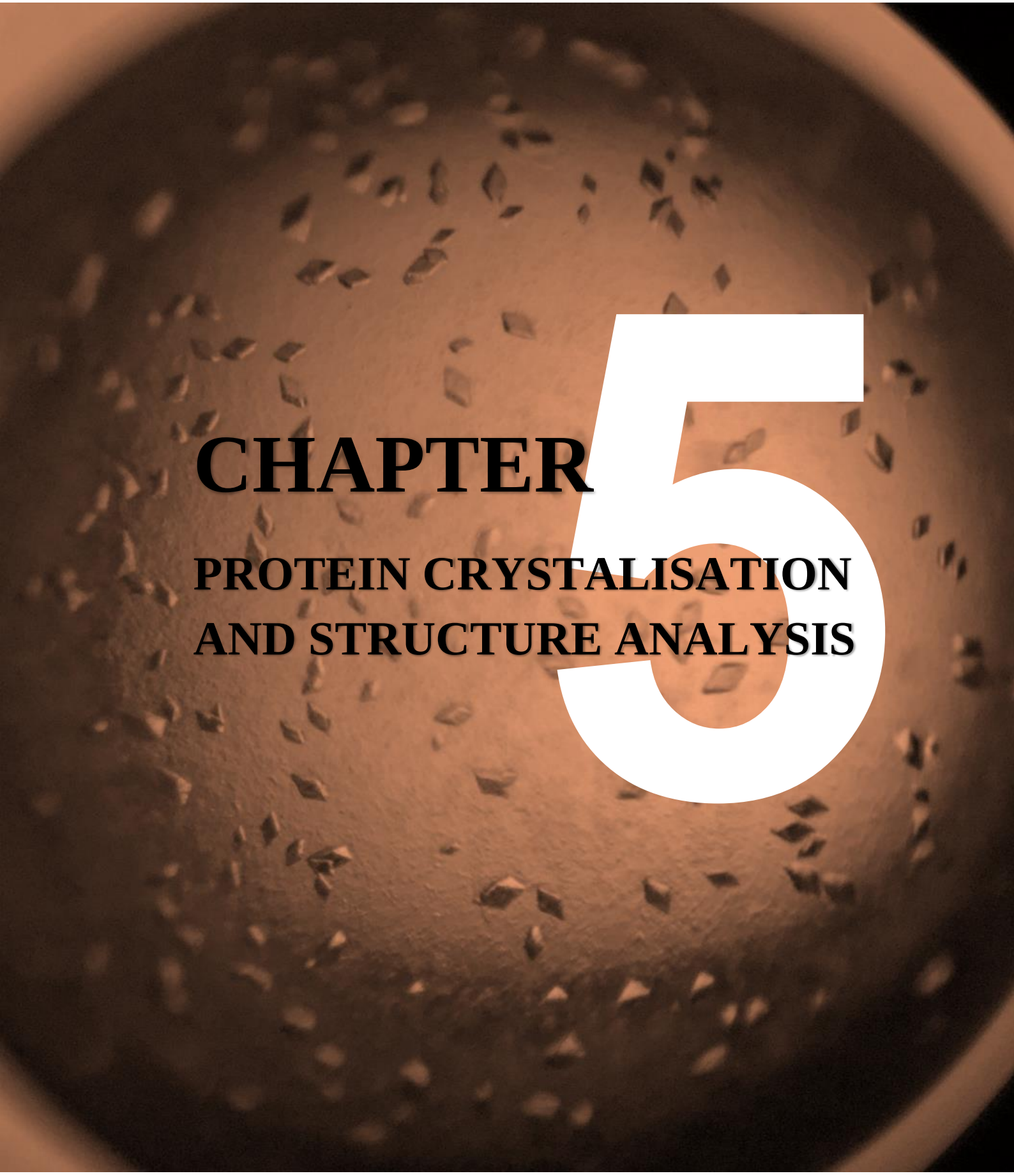
4.3 Conclusions

In this chapter, a study of the six structural subunits of the *C. phytofermentans* bacterial microcompartment shell has been reported through the coordinated recombinant expression of the respective genes in *E. coli*. The effect of the production of single shell proteins and in combination with one another has been looked at through the use of electron microscopy. Specifically, evidence has been sought for the formation of higher order structures such as protein sheets, filaments or compartments after embedding, thin sectioning and TEM of cells producing these proteins.

Different internal cytoplasmic structures were observed by the overproduction of individual shell proteins, demonstrating not only that the proteins were being produced but that they could form higher order structures. The main structures observed through this imaging approach were filaments and sheets, although a number of aggregates and other shapes/deposits were also observed. Such images have been reported before with the overproduction of individual shell proteins associated with both metabolosomes and carboxysomes.

Attempts were made to form empty microcompartments by coproducing the six shell proteins with different induction times and with different levels of inducer. In this case, although a range of different structures were observed within the cell, none-could really be identified as intact BMCs. However, on purification, some structures were observed that looked a bit like BMCs although the structures had somewhat soft edges. Nonetheless, these structures deserve further investigation.

A number of other shell-gene combinations were also investigated. These, too, made a range of interesting cytoplasmic geometric structures although none of the combinations appeared to form discrete BMCs. It has to be pointed out that in terms of metabolosomes, there is still only one report of recombinant BMCs being produced within cells, and that is the *C. freundii* Pdu system (Parsons *et al.*, 2008).



CHAPTER 5

PROTEIN CRYSTALISATION AND STRUCTURE ANALYSIS

5.1 Introduction

C. phytofermentans has six shell proteins (Cphy_1176, 1180, 1181, 1182, 1184, 1186) that assemble together in a higher level of organisation to generate a polyhedral BMC. Individually, the various shell proteins form hexameric- or pentameric-shaped tiles. These tiles represent either a 6-fold, 5-fold or 3-fold, and level of symmetry, generated from a hexameric, pentameric or trimeric shell protein.

The relationship between protein structure and function is critical in the understanding of how processes work at the molecular level, and this is especially true with macromolecular complexes which are composed of many different subunits. BMCs represent one of the largest protein complexes found in prokaryotic cells and the shell-proteins that make up the outer casing play an essential role in acting as a selective permeability barrier. In *C. phytofermentans* there are six shell proteins that must have different properties, characteristics that can really only be realised from a detailed structure determination of each individual unit. In order to appreciate the contribution that each shell protein makes to the overall BMC, the individual shell proteins were subject to crystallisation in order to establish if they were suitable for structured determination by X-ray crystallography.

The aims of this chapter were to determine the crystal structures of three shell proteins, Cphy_1181, Cphy_1182 and Cphy_1186. Although some structures are known for shell proteins, for each specific BMC it is envisaged that there will be differences to make the BMC specific for its host cell and for the particular metabolic process for which it has evolved. Moreover, the

structure determination of several shell proteins should help in understanding how the individual tiles interact to give the overall outer structure.

5.2 Results

As described in Chapter 3, the genes encoding the shell proteins Cphy_1181, Cphy_1182 and Cphy_1186 were cloned into pET14b, thereby allowing the shell proteins to be overproduced recombinantly in *E. coli* with an N-terminal poly-His extension. The cells were harvested by centrifugation before purification by immobilised metal affinity chromatography. The purified proteins were then subject to crystallisation trials using a number of standard crystallisation screens and proteins that crystallised were then subject to diffraction studies.

5.2.1 Structure analysis of Cphy_1181

In comparison to the *C. freundii* Pdu shell proteins, the BMC domain on Cphy_1181 shares the highest level of similarity to PduK, which interestingly is a shell protein for which no structure has been determined. A comparison is displayed between PduK and Cphy_1181 in **Figure 4.5**. PduK has 160 amino acids including a C-terminal extension of ~70 amino acids, which includes four cysteine residues C_{132}, C_{135}, C_{140} & C_{151} and was thought to possibly form an Fe-S redox centre but for which no evidence has been obtained (Chun *et al.*, 2014). In contrast, Cphy_1181 is missing a C-terminal extension but contains a smaller ~ 70 amino acid N-terminal extension, which contains just one cysteine (C_{24}) and is unlikely to be part of an Fe-S

centre **(Figure 5.1)**. As described previously, the gene encoding Cphy_1181 was cloned into pET14b (hexa histidine-tag) and transformed into *E. coli* cells. For more details about the overproduction and purification of Cphy_1181 protein see **Chapter 3**.

```
         10          20          30          40          50          60
MGKSLGFIEI SGVTAAIDAL DIMCKTAGVE LVTWERKMGG RLVTIIVSGS VSEVTQAVES

         70          80          90         100
AVNQAIKKPV AHAVIANPHE EVLRLVDISA SRMKPQNEEM
```

Figure 5.1: Cphy_1181 shell protein 100 amino acid sequence.

5.2.1.1 Crystallisation of Cphy_1181

Two litres of culture that had been grown overnight after induction with IPTG were harvested by centrifugation and the protein extracted and purified by immobilised metal affinity chromatography. Half of the purified protein was treated with thrombin in order to cleave off the hexa-His tag. The purified protein samples were concentrated and applied to the crystallisation screens. A number of different concentrations of the isolated protein were made as shown in **Table 5.1**.

Table 5.1: Protein concentrations used in crystallography screening to optimise the best condition. Those concentrations were used as starting point to find the required concentration for making nice crystal.

Protein status	Protein concentrations used for screening		
Cphy_1181-Hexa-His-Tag	8 mg/ mL		
Cphy_1181	4 mg/ mL	8 mg/ mL	14.3 mg/ mL

The crystallisation trials were initiated by taking 1 µL of each concentration of Cphy_1181 and mixing it directly with 1 µL of the corresponding reservoir of Molecular Dimensions Structure Screen I and II using the hanging drop vapour diffusion method in 24 well Linbro plates. From this screen just one drop (The protein with no His-tag) from the 50 different conditions was observed to contain a crystal after a week of equilibration. However, this reservoir contains 30 % propanol, which is volatile making the drop difficult to handle. Thus, it was decided to look at different conditions that crystallised a homologous protein associated with carboxysome formation, CcmK4. The CcmK4 protein crystallised in conditions that included 0.1 M $MgCl_2$, 0.1 M sodium citrate pH 5.0, with 15 % (w/v) PEG 4000 (Cai *et al.*, 2014). Thus 120 conditions around these conditions were screened in order to identify those that encouraged the crystallisation of Cphy_1181. After a week around 30 % of the conditions were found to have produced crystals. From this screen the best crystallisation reservoir conditions were selected **(Figure 5.2 and 5.3)**.

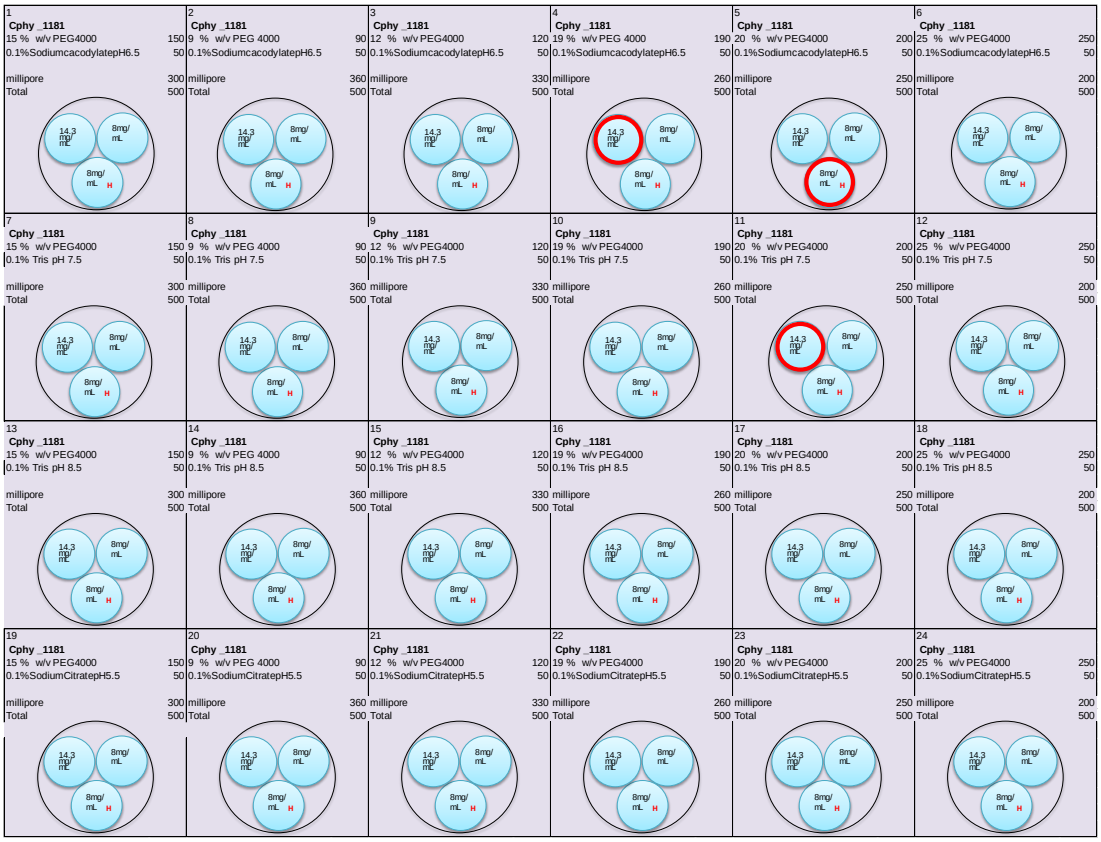


Figure 5.2: Optimisation of Cphy_1181 crystallisation conditions. The conditions were refined via varying the concentration of PEG 4000 horizontally and the pH of 0.1 % Sodium cacodylate, Tris & Sodium citrate vertically on a 24-well crystallisation plate. Conditions that produced good crystals are circled with red.

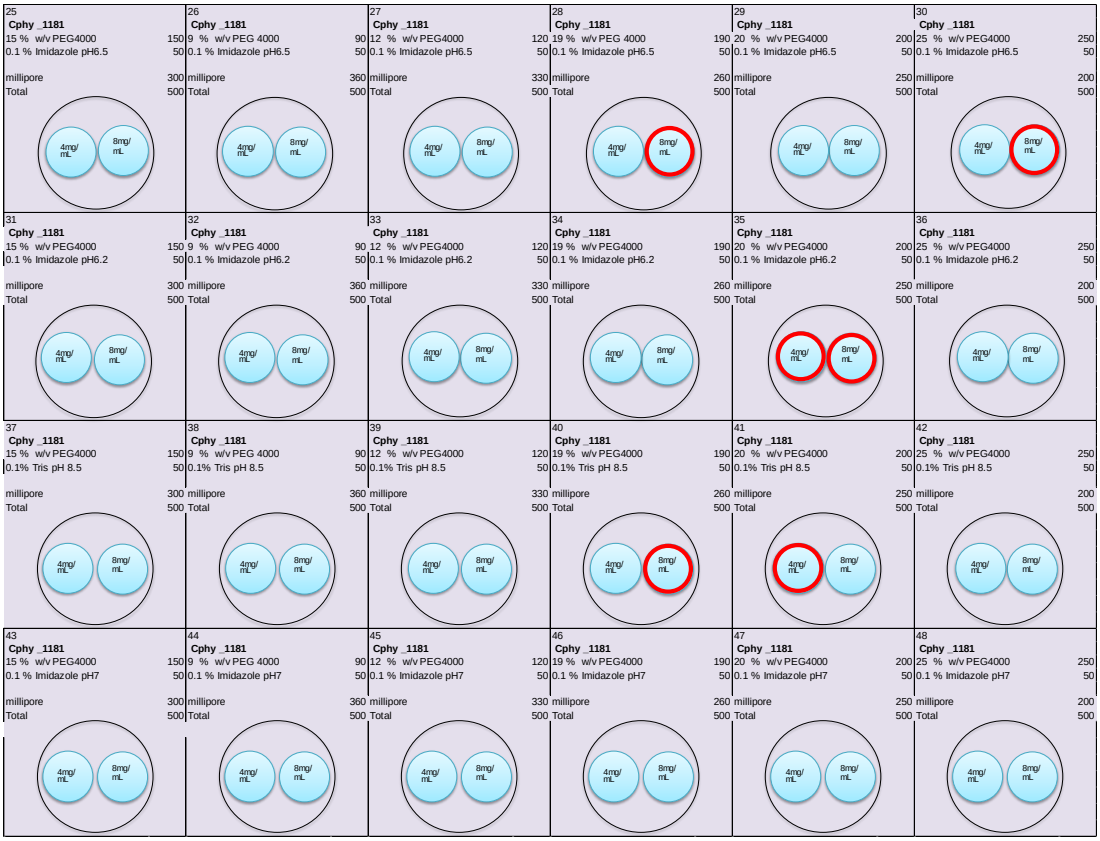


Figure 5.3: Optimisation of Cphy_1181 crystallisation conditions. The conditions were refined via varying the concentration of PEG 4000 horizontally and the pH of 0.1 % Imidazole vertically on a 24-well crystallisation plate. Conditions that produced good crystals are circled with red.

The best crystals were observed in reservoir number 35 (20 % W/V PEG 4000, 0.1 % Imidazole pH 6.2), which produced diamond plate crystals **(Figure 5.4)**. A number of crystals were selected and subject to X-ray diffraction, which exhibited a characteristic protein diffraction pattern. Unfortunately, the data that was collected did not process properly and hence it was not possible to solve the structure. This work was not followed up, but it would be quite straightforward to regrow the crystals using the optimised conditions and investigate this further.

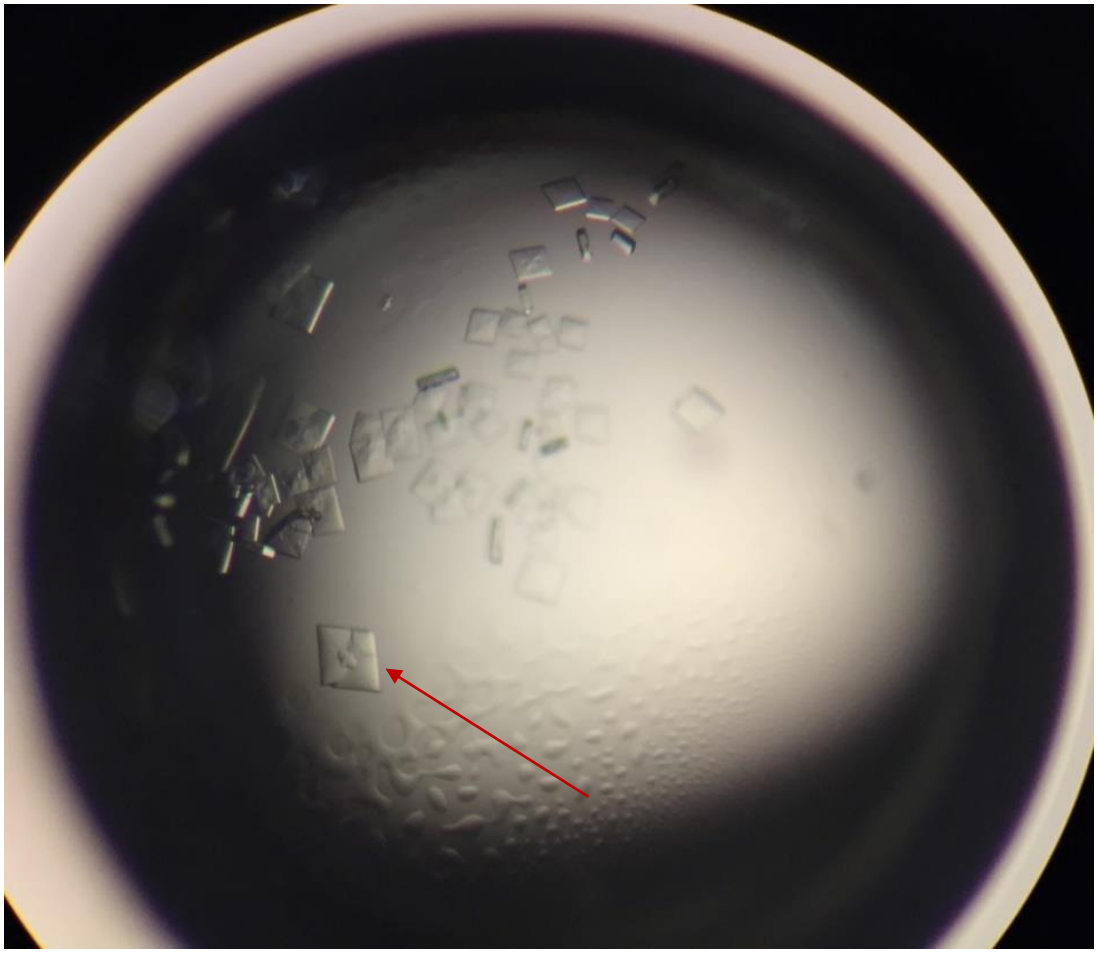


Figure 5.4: Hanging drop of Cphy_1181 (PduK-like) protein crystals. The reservoir here is number 35 (Figure 5.3), which contained 20 % W/V PEG 4000/ 0.1 % Imidazole pH 6.2 and protein concentration was 4 mg/ mL with no His-tag.

5.2.2 Structure analysis of Cphy_1182

An initial characterisation of Cphy_1182 revealed that the protein is soluble and forms higher order tubes structures of about 25 nm in diameter **(Section 4.2.1.4)**. The presence of these structures suggests that Cphy_1182 may form ordered structures in crystallisation conditions. A sequence analysis of Cphy_1182 reveals that the protein has a single-BMC-domain and has an N-terminal extension of ~ 75 amino acids. The BMC domain displays a high level of similarity to the *S. enterica* PduA (Protein Data Bank accession code

3NGK) (Pang *et al.*, 2014). The PduA hexamer has a central pore opening along its axis, that suggestive of 1,2-PD transport (Chun *et al.*, 2014). In this Chapter, the crystal structure of Cphy_1182 was solved by molecular replacement using PduA as the search model. However, the N-terminal extension was not observed suggesting that this region of the protein may have been cleaved.

5.2.2.1 Crystallization of Cphy_1182

As mentioned previously, there are no Trp, Tyr or Cys residues within the primary structure of Cphy_1182, and hence it was not possible to accurately measure the concentration of the protein by UV absorbance and thus the protein concentration was roughly estimated from SDS gels of the purified material. Purified protein was entered into Molecular Dimensions Structure Screen I and II using the hanging drop vapour diffusion method in 24 well Linbro plates. One condition from this screen was observed to contain crystals after a week of equilibration **(Figure 5.5)**. The successful condition was from the Structure Screen I/reservoir 19 (0.2 M Zinc acetate dihydrate, 0.1 M Sodium HEPES pH6.5, and 18 % w/v PEG 8000). The crystals of Cphy_1182 presented as rectangular rods.

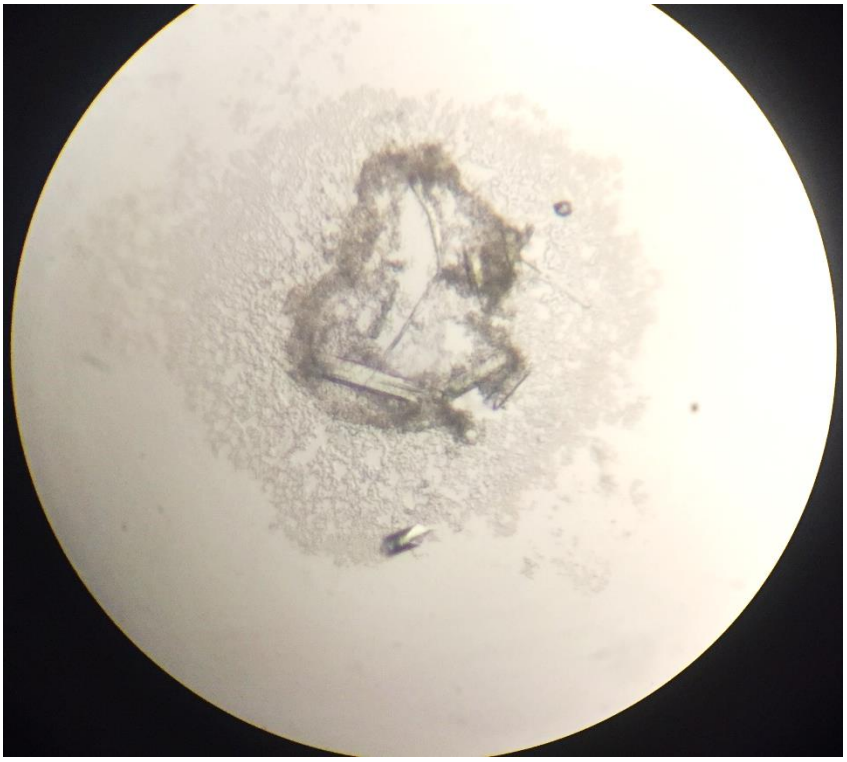


Figure 5.5: Hanging drop of Cphy_1182 protein showing different sizes of crystals. The reservoir is number 19 of Structure Screen I that contain 0.2 M Zinc acetate dihydrate, 0.1 M Sodium HEPES pH6.5, and 18 % w/v PEG 8000.

5.2.2.2 Data collection and molecular replacement

Crystals were mounted and collected at the DIAMOND synchrotron source on beamline I03. The crystals were found to have a space group of $P2_12_12_1$ and diffracted to a resolution of 2.2 Å with 97 % complete data. The structure was solved by molecular replacement using PduA (PDB ID: 3ngk) as a search model. MOLREP found two molecules in the AU with an initial R factor of 55 % and an R_{free} of 56 %. Thirty rounds of REFMAC reduced the R factor to 30 and R_{free} to 34 %. Further manual modelling, which was validated and assessed via Ramachandran plot building, coupled with further

refinement resulted in a final R factor of 22 % and an R_{free} of 27 % **(Figure 5.6)**. The structure determination was performed Professor Dave Brown (University of Kent).

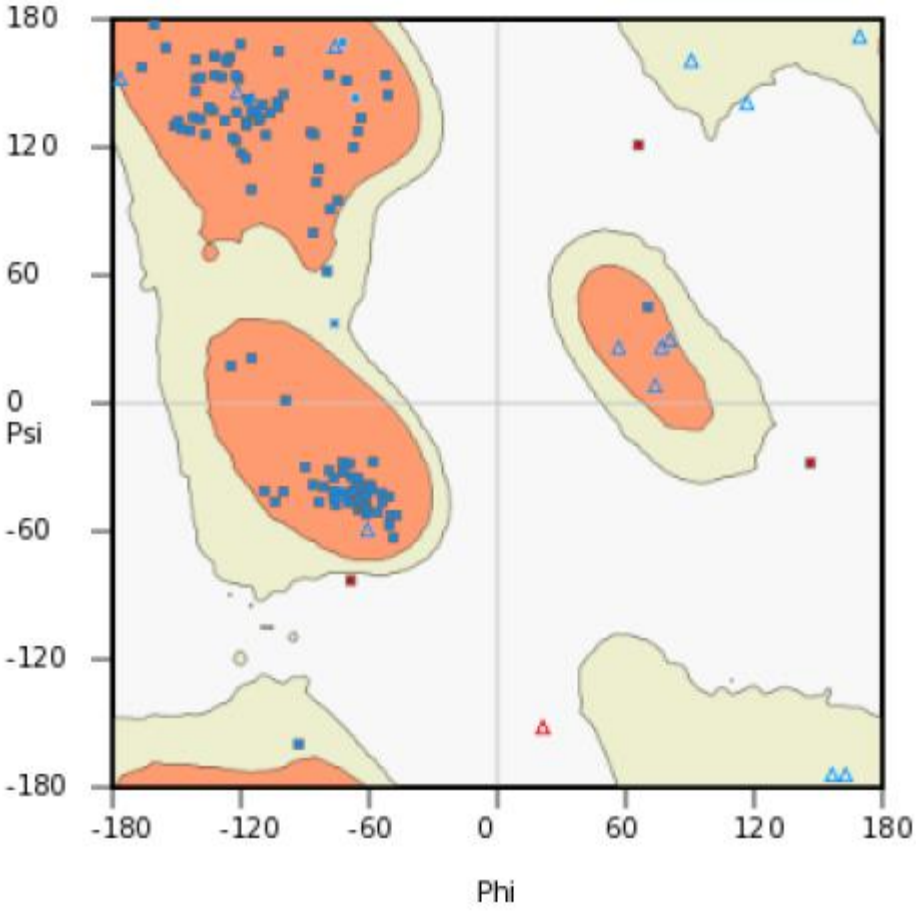


Figure 5.6: A Ramachandran plot generated from *C. phytofermentans* Cphy_1182, a dimer protein that contains both β-sheet and α-helix.

Strangely, the Cphy_1182 subunits did not assemble into a hexamer as is usually seen with single-BMC-domain structures. Rather, the protein was found to crystallise as a dimer and this is what makes this structure particularly interesting **(Figure 5.7)**.

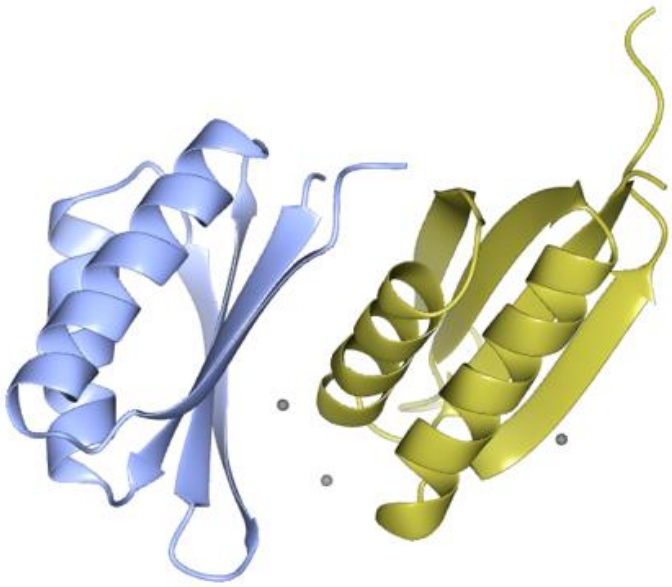


Figure 5.7: Cartoon representation of the tertiary structure of Cphy_1182 dimer which comprises four Zn. This figure was produced using PyMOL (DeLano and Lam, 2005).

The model was composed of 156 residues in two copies of the protein (chain A residues 83-158 and chain B residues 82-161) with 9 waters and four zinc atoms, which may help promote the formation of the dimer. The likely biologically relevant hexamer was generated by overlaying molecule A of Cphy_1182 with a symmetry generated hexamer of PduA (PDB ID: 3ngk). The two BMC domains superimpose with 80 aligned residues. As with other

hexameric shell proteins, one face of the manually produced hexamer is concave and the other convex **(Figure 5.8)**.

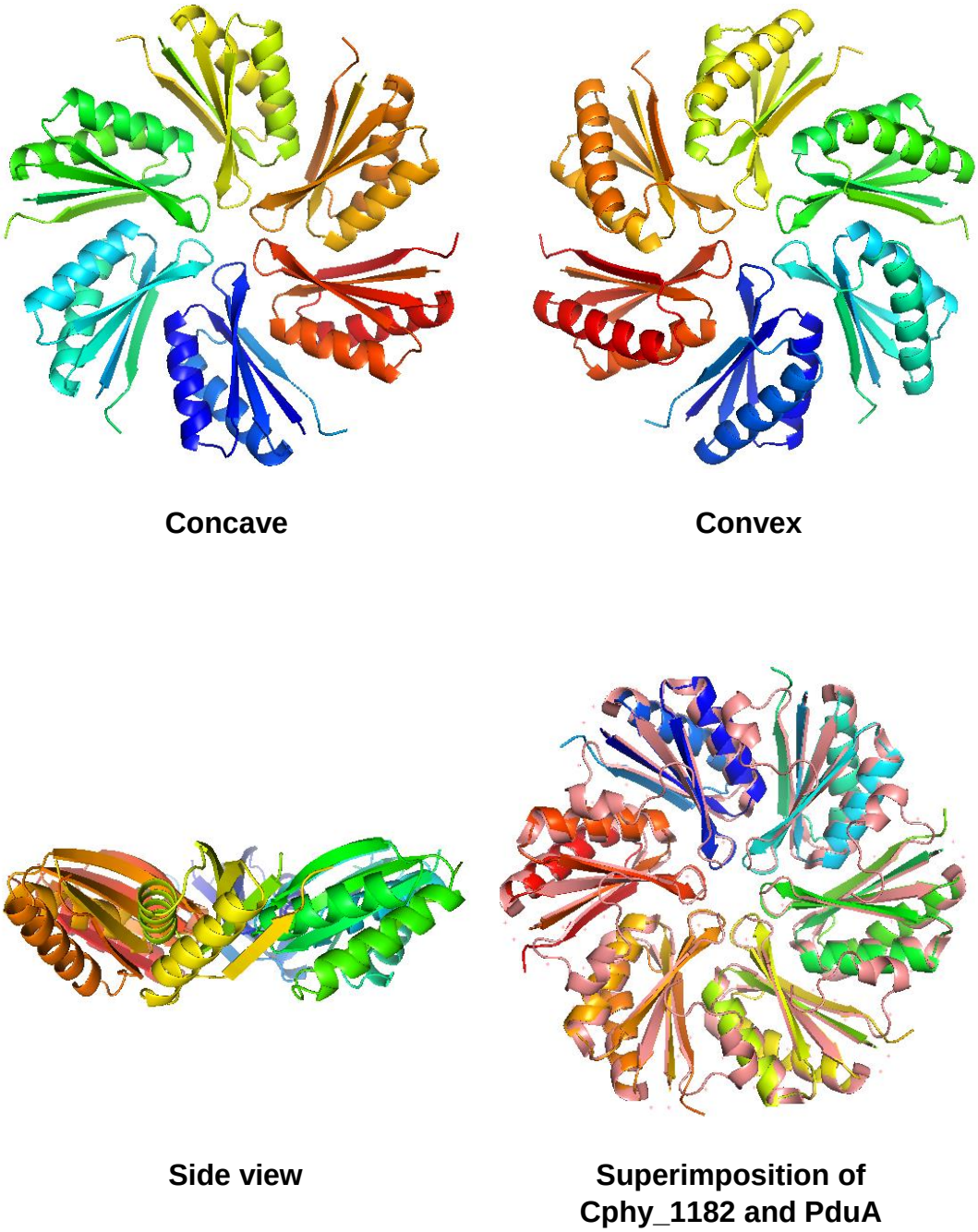


Figure 5.8: Surface representation of the generated hexamer of Cphy_1182. On the top right is a convex face of the hexamer and on the top left is a concave face of the hexamer. At the bottom right is a superimposition of Cphy_1182 and PduA. At the bottom left is a side view of the hexamer where the convex at the top and concave down. This figure was produced using PyMOL (DeLano and Lam, 2005).

The superimposition shows that Cphy_1182 has a similar size central pore (~6 Å) to that found in PduA, which suggests that the Cphy_1182 shell protein may also play a role in the movement of 1,2-PD. The packing of Cphy_1182 within the crystal lattice reveals a higher order packing, however, there was no evidence of crystallographic hexamers **(Figure 5.9)**. This is in contrast to the crystallographically observed PduA hexamers,

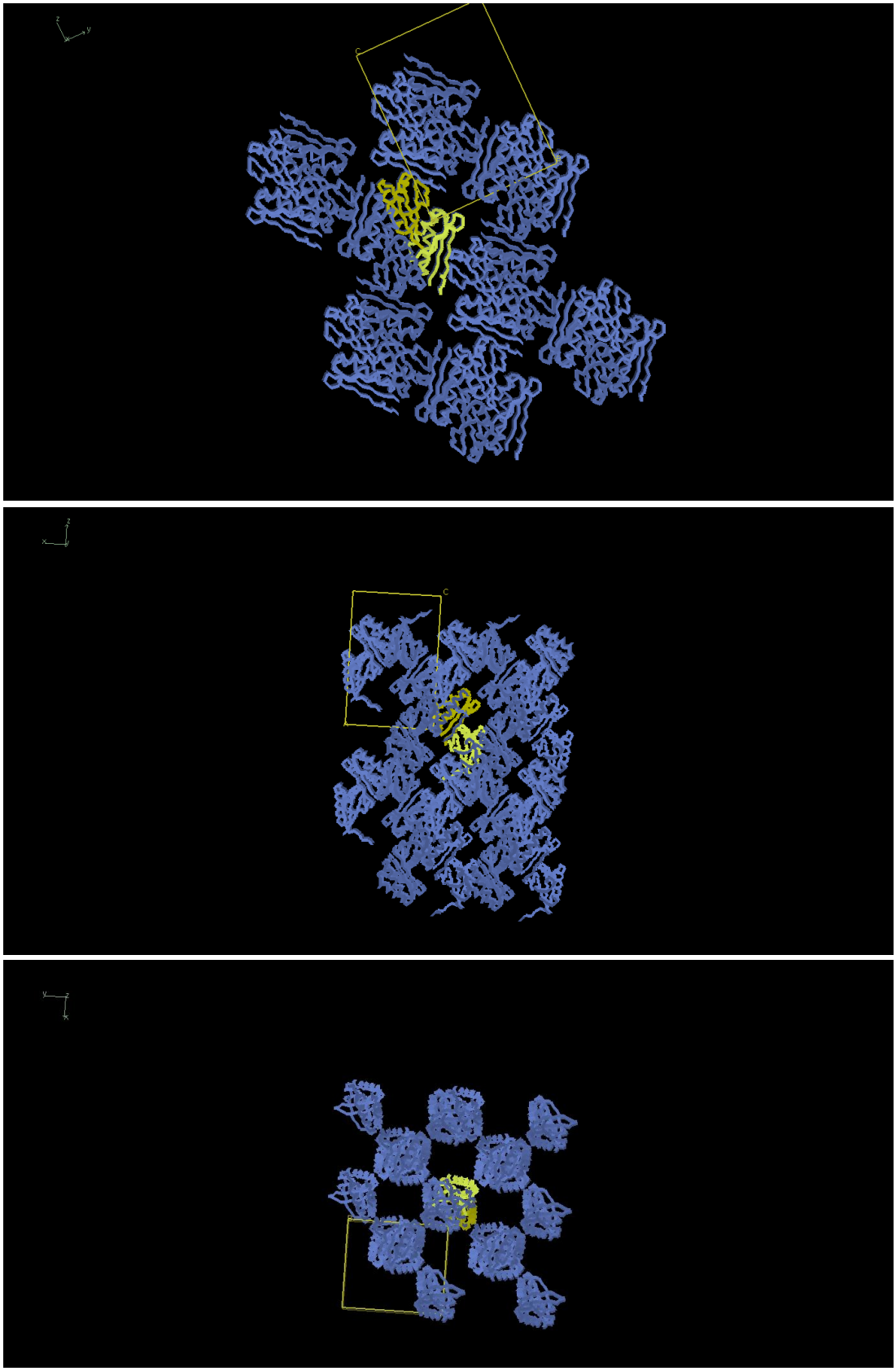


Figure 5.9: Crystal packing of Cphy_1182 dimers. One dimer is shaded with yellow. This figure was produced using PyMOL (DeLano and Lam, 2005).

which pack into uniform molecular sheets (Yeates *et al.*, 2010). The zinc atoms that are found within the Cphy_1182 dimer are ligated by a histidine and two glutamic acid residues **(Figure 5.10)**. This arrangement suggests that the zinc may help to maintain the dimeric Cphy_1182.

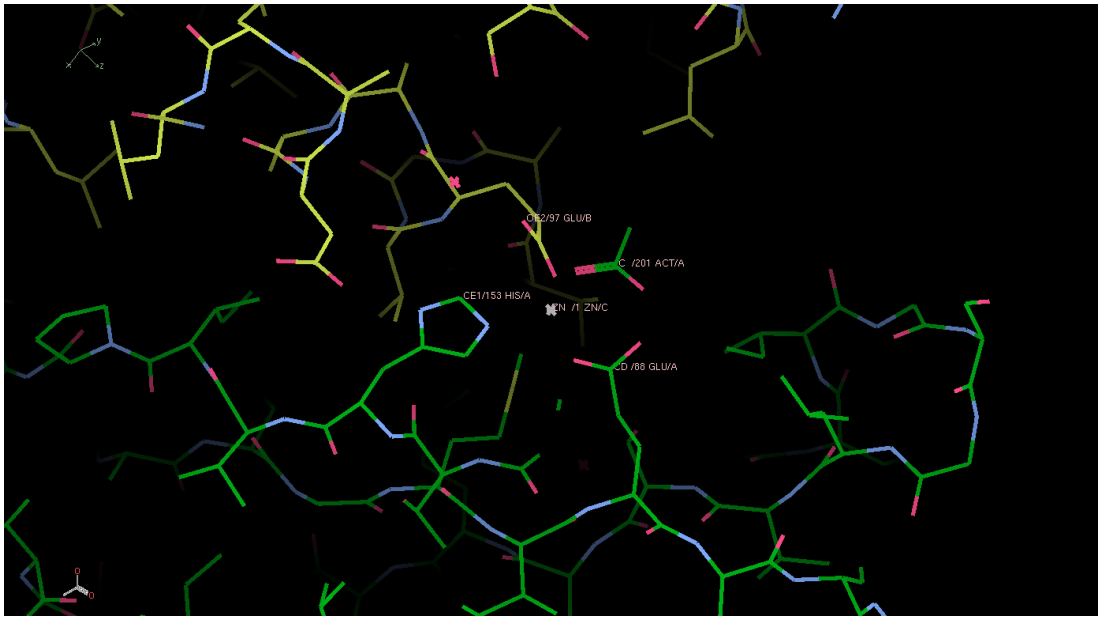


Figure 5.10: A model of the zinc atom interact a Histidine and two Glutamic residues. This figure was produced using PyMOL (DeLano and Lam, 2005).

Within PduA it has been shown that three residues, Arg-76, Lys-26 and Val-51, from two subunits of the PduA hexamer are responsible of hexamer-hexamer interactions, to help in the tiling of the hexamer into a sheet (Pang *et al.*, 2014). The edge of the Cphy_1182 subunit also have an arginine residue, Arg-157, which may play an important role in the hexamer-hexamer interaction. Furthermore, sequence alignment indicates that Arg-157 of Cphy_1182 is located in the same position of the BMC domain as Arg-76 of

PduA **(Figure 4.3)**. Three other residues Pro-158, Asp-128 and Asn-107 are also well-placed to assist in such interactions **(Figure 5.11)** and **(Figure 5.12)**.

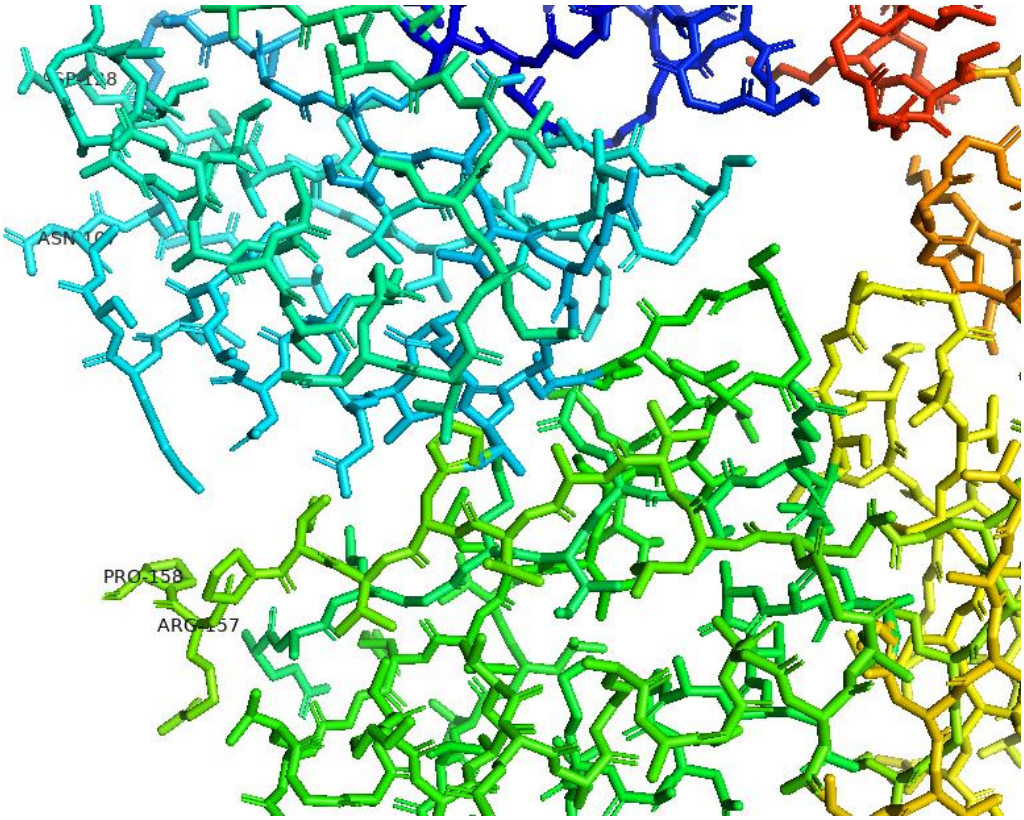


Figure 5.11: Two Cphy_1182 subunits showing the edge residues Arg-157, Pro-158, Asp-128 and Asn-107. This figure was produced using PyMOL (DeLano and Lam, 2005).

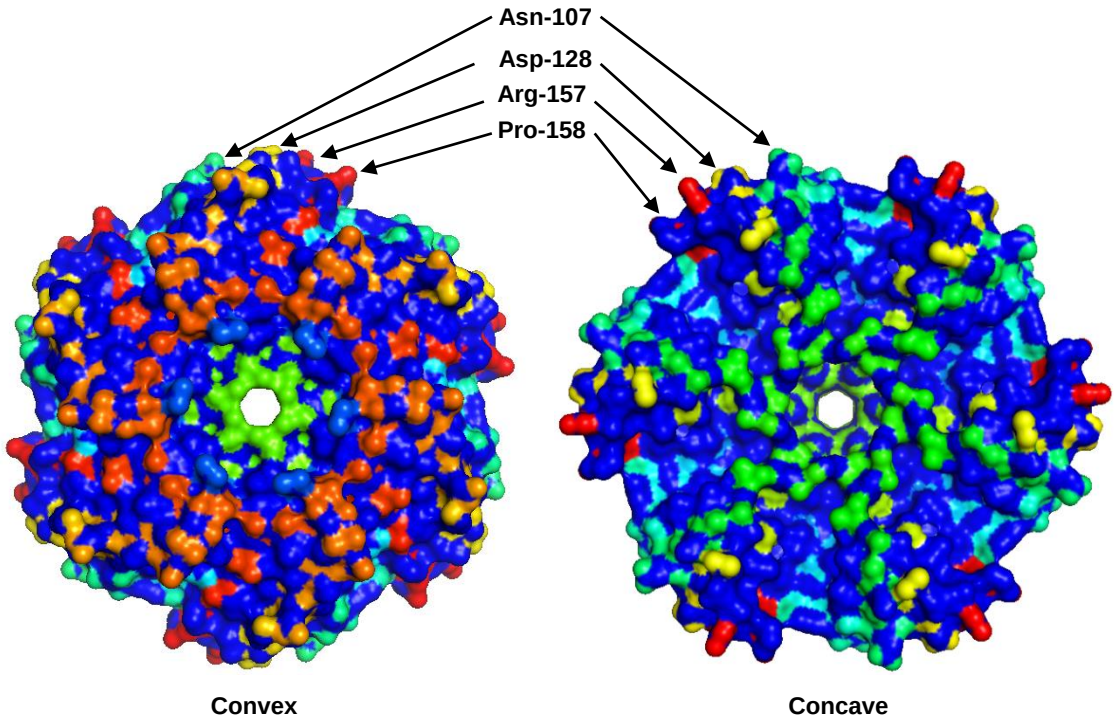


Figure 5.12: A surface representation of Cphy_1182 hexamer showing the sides and the edge residues of the hexamer. This figure was produced using PyMOL (DeLano and Lam, 2005).

As with all BMC-domain shell proteins, the hexameric tile displays a central pore on the six-fold axis of the structure. With Cphy_1182 this pore is lined with six serine residues at positions 118 of each subunit **(Figure 5.13)**. The

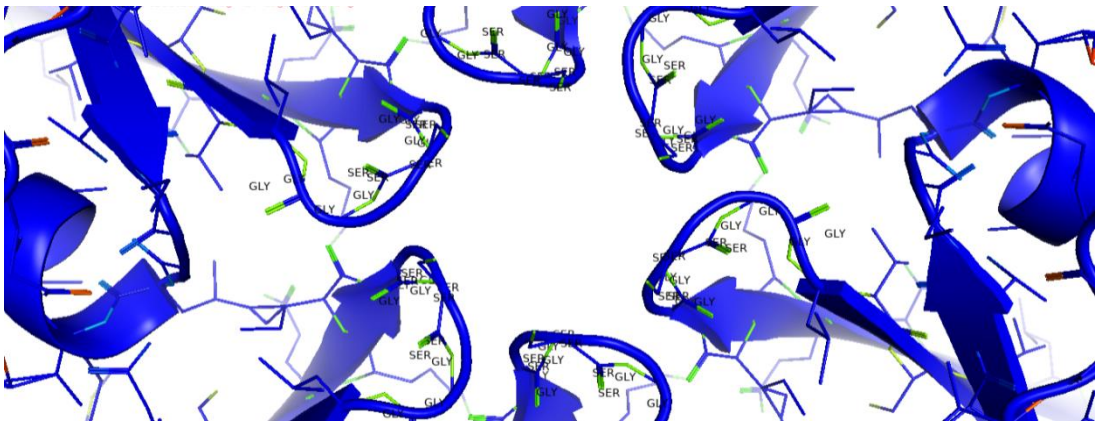


Figure 5.13: The subunit pore of Cphy_1182 showing the positions of residues that line the pore. This figure was produced using PyMOL (DeLano and Lam, 2005).

serine residue helps make the pore more polar in character.

In conclusion, the crystal structure of Cphy_1182 reveals that the shell protein crystallises as a dimer rather than a hexamer, possibly due to the interaction of the indicial subunits with zinc. The first 75 residues of the N-terminal extension of the protein are not visible within the structure, either because they have been cleaved off or because the N-terminus has a high degree of flexibility. Six subunits of Cphy_1182 were used to generate a hexameric model to represent the quaternary structure of the protein, a structure that likely reflects its physiological state as a tile within the BMC. The structure of Cphy_1182 has identified some of the residues on the edge of the hexamer that may play important roles in hexamer-hexamer interactions. The structure also identified residues on the β-hairpin that is located in the central pore and which may play a role in allowing substrates/products to move into/out of the BMC.

5.2.3 Structure analysis of Cphy_1186

Sequence analysis of Cphy_1186 suggested that the protein shares some similarity with PduT. When the protein was purified it was found to have a red/brown colour, consistent with the presence of a redox group such as an Fe-S centre. Indeed, a UV-visible spectrum gave a spectrum with absorption of around 420 nm, again consistent with the presence of an Fe-S system. The presence of an 4Fe-4S cluster suggests that Cphy_1186 may serve as a conduit for electron transfer. In addition, the sequence of the shell protein highlighted that the subunit contains just two cysteines (C_{38} and C_{133}) **(Table 5.2 and Figure 5.14)**.

Table 5.2: Cphy_1186 sequence analysis for cysteine detection.

Name	Accession	Location	Relative to	HGVS Name	Sequence
1Cys	ABX41564.1	38	Seq start	ABX41564.1:p.38	DIDIVEAQTV**C**PGKYIILITG
2Cys	ABX41564.1	133	Seq start	ABX41564.1:p.133	LVEVARGRVM**C**GKSYLMLTG

```
         10         20         30         40         50         60
MSKAIGMVEY KTVSSGIMAA DLMVKTADID IVEAQTVCPG KYIILITGDL SAVNASVEAA

         70         80         90        100        110        120
RIQFETHLID SFILGNPHDG ILPAIYGASV VEEIEALGVL ETYSAASIIV AADVAAKTAA

        130        140        150        160        170        180
VELVEVRVAR GMCGKSYLML TGEIASVTAS IEAAKKAIGE NGMYLDSSVL AHPDKKLRNK


LM
```

Figure 5.14: Cphy_1186 shell protein 182 amino acid sequence.

EPR was undertaken not only to confirm the presence of the iron sulphur cluster, but also to detect the number of clusters. When the cysteine residues (C_{38} and C_{133}) are mutated either collectively or individually to alanine residues (Section 5.2.3.8), the EPR spectrum corresponding to the 4Fe-4S cluster is lost or altered (Section 5.2.3.9), indicating that both cysteines are crucial for coordination of the cluster. In this section, the determination of the structure of Cphy_1186 is described and reveals that this shell protein is trimer with a binding site for 4Fe-4S cluster.

5.2.3.1 Anaerobic purification of Cphy_1186

Aerobic overexpression and isolation of Cphy_1186 resulted in a brown-coloured protein in which the UV-visible spectrum had a broad absorption maximum at 420 nm **(Section 3.2.3.6)**, suggesting the presence of iron-sulfur cluster. Cphy_1186 was purified anaerobically from three litres of recombinant culture via IMAC and desalted on a PD-10 column in a Belle Technology glove-box under an atmosphere of nitrogen. The anaerobic environment helps to preserve Fe-S centres which are often sensitive to oxygen **(Figure 5.15)**. SDS PAGE was used to analyse the darkest coloured fractions of the purified Cphy_1186 protein that eluted form the IMAC column **(Figure 5.16)**. The isolated fraction was concentrated to 20 mg/mL (> 250 µM). A sample was subject to UV-vis spectroscopy in the presence and absence of dithionite **(Figure 5.16)**.

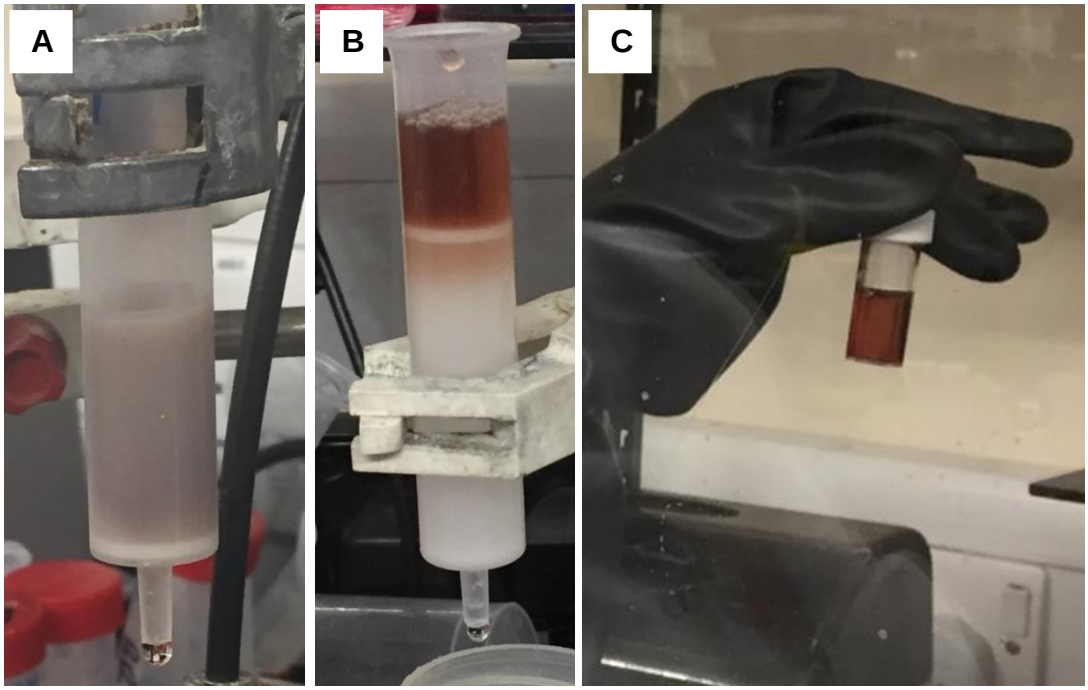


Figure 5.15: Anaerobic purification technique of Cphy_1186. A is the Cphy_1186 soluble protein loaded into the Ni column. B is the distinct colour of Cphy_1186 applied into PD-10 column. C is Cphy_1186 eluted from the Nickel column.

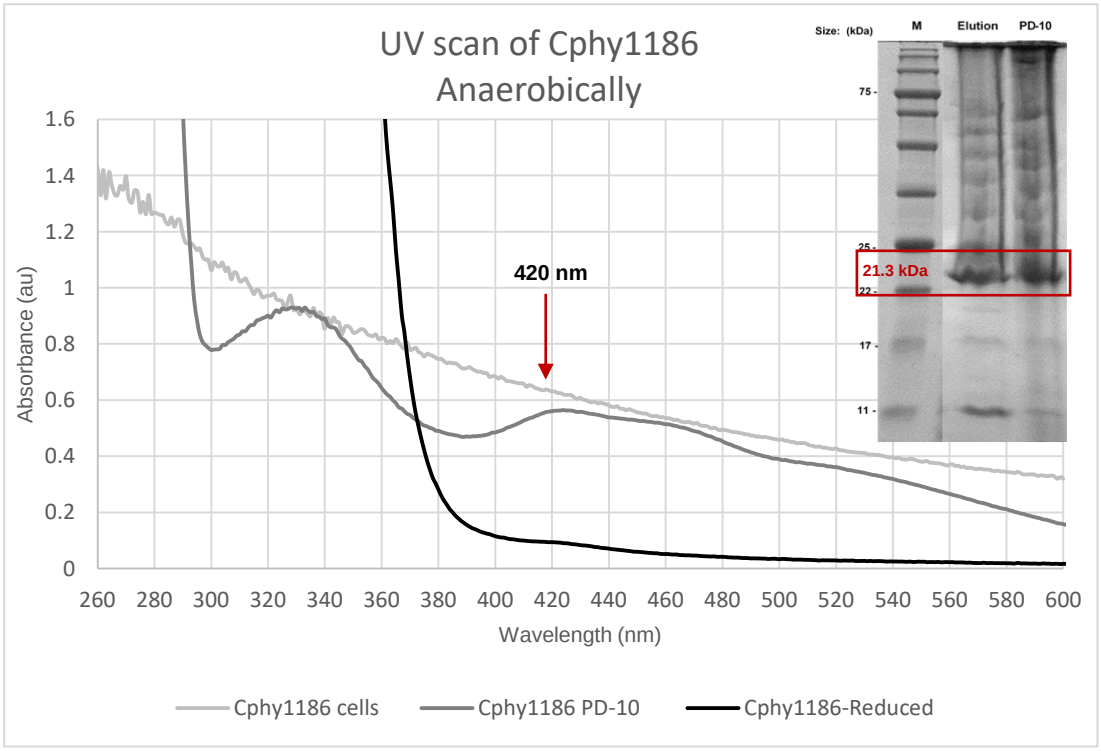


Figure 5.16: UV-visible absorbance spectrum of anaerobic isolated Hexa-His tagged Cphy_1186 protein. The UV-visible spectrum of purified Cphy_1186 proteins in 20 mM Tris-HCl (pH 8.0), with 100 mM NaCl. Light grey line represents the whole cells of Cphy_1186. Medium grey line represents the purified Cphy_1186 and the absorbance concentrated at 420 nm represents the Fe–S cluster. The black line represent the absorbance spectrum of purified Cphy_1186 after reduction of the cluster with sodium dithionite to the [Fe–S]+ form. On the top right is SDS-PAGE of anaerobic IMAC and PD-10 purification of Cphy_1186 protein.

A total of four samples were prepared for EPR analysis. The four samples were composed of purified Cphy_1186, reduced purified Cphy_1186 protein, whole cells expressing Cphy_1186 and reduced whole cells expressing Cphy_1186. The samples were pipetted into EPR tubes and were stored frozen in liquid nitrogen for analysis. In addition, a quantity of the anaerobic purified Cphy_1186 protein was used for anaerobic crystallography.

5.2.3.2 Crystallization of Cphy_1186

The isolated protein was treated with thrombin before crystallisation in order to cleave the hexa-histidine tag. Trials were initiated by taking 1 µL of 20 mg/ mL of anaerobically purified Cphy_1186 and mixing with 1 µL of the corresponding reservoir of Molecular Dimensions Structure Screen I and II using the hanging drop method. The hanging drop was placed on a glass cover slip that was placed over the well, and incubated at 18 °C. Microcrystals were observed in wells that contained PEG after a week of equilibration. A few of these conditions were selected for optimisation. Finally, it was found that Cphy_1186 crystallised best in conditions with PEG 8000. By using these conditions with a range of protein concentrations, 7, 10 and 20 mg/ mL, many crystals were formed in the various drops, but the best diffracting crystal was grown in reservoir number 9 (18 % W/V PEG 8000, 0.1 % Tris pH 7.5, 0.2 M Sodium acetate with a Cphy_1186 concentration of 20 mg/ mL) **(Figure 5.17)**. The crystals were light brown in colour and had shapes that varied between triangular and diamond **(Figure 5.18)**.

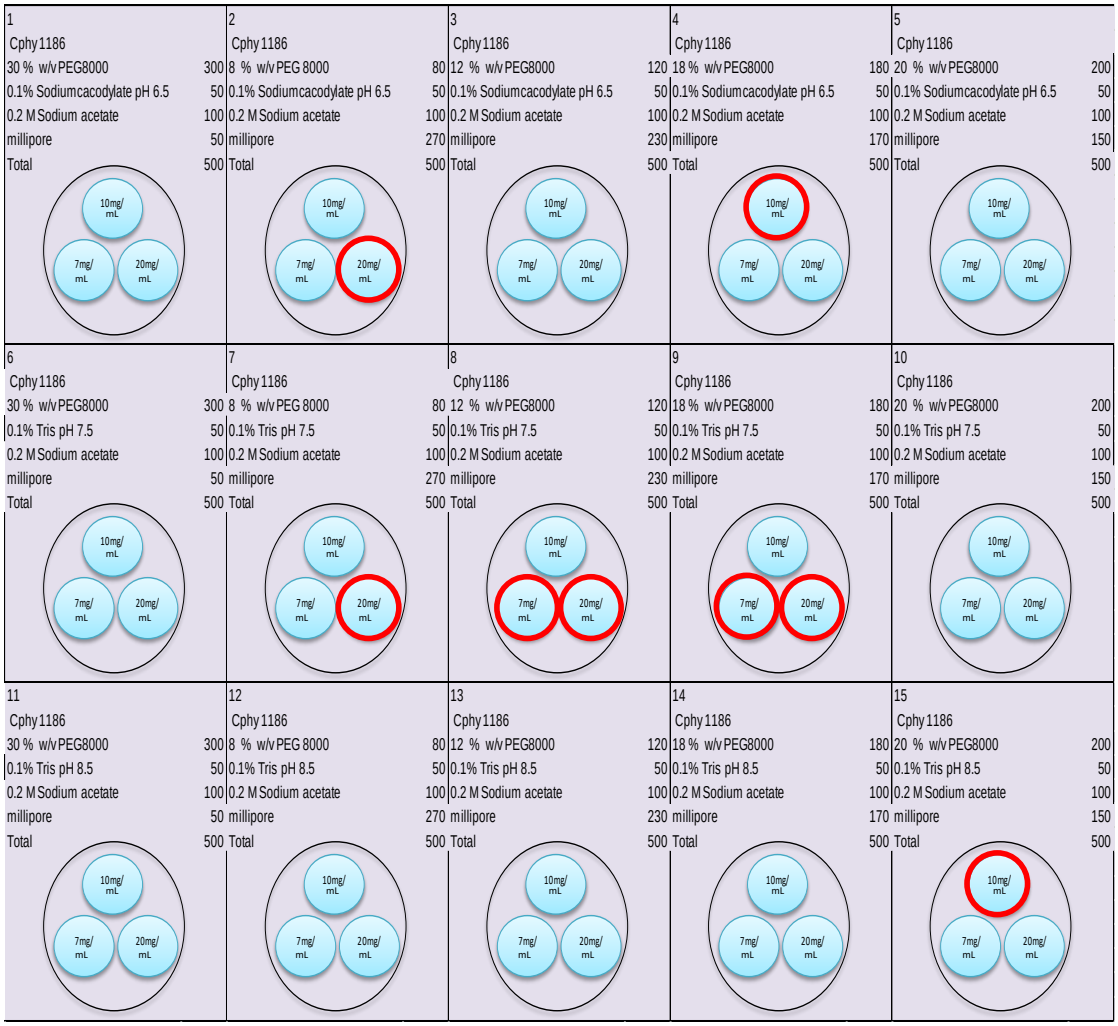


Figure 5.17: Optimised conditions of Cphy_1186 anaerobic crystallisation method. The conditions were refined via varying the concentration of PEG 8000 horizontally and the pH of 0.1 % Sodium cacodylate, Tris & Sodium citrate vertically on a 15-well crystallisation plate. Protein concentration are 7, 10 and 20 mg/ mL. Reservoirs that produced good crystals are circled with red.

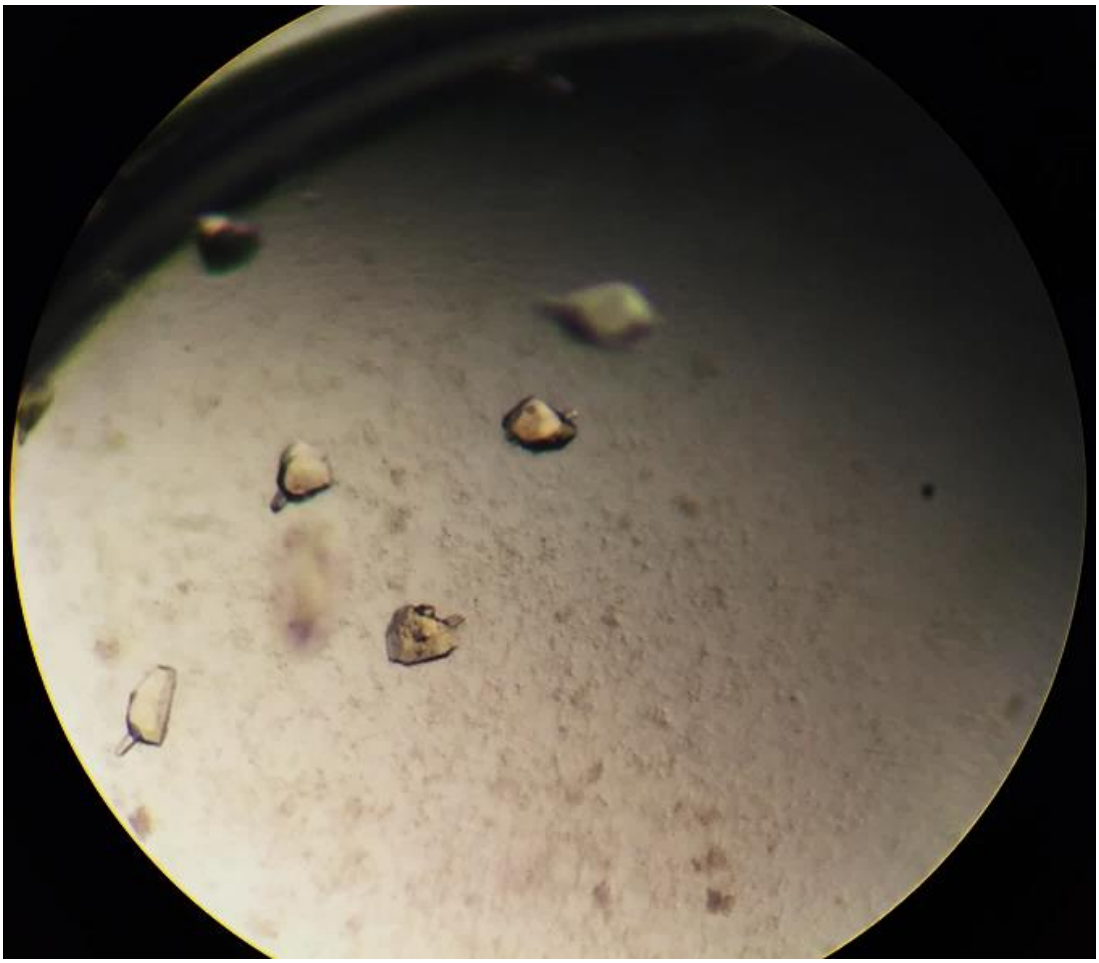


Figure 5.18: Anaerobic hanging drop of Cphy_1186 (PduT-like) protein crystal. The reservoir here is number 9 (Figure 5.17), which contained 18 % W/V PEG 8000/ 0.1 % Tris pH 7.5, 0.2 M Sodium acetate and protein concentration was 20 mg/ mL.

5.2.3.3 Data collection

Crystals were picked with a 0.1 mm litholoop and immediately soaked in a cryoprotectant solution and stored in liquid nitrogen **(Figure 5.19A)**. The cryoprotectant solution was made of 80 % of the crystal growth reservoir solution mixed with 20 % glycerol, which was added in order to flash freeze with liquid nitrogen. The crystal was subject to illumination with synchrotron X-ray radiation at beamline I04-1 at the Diamond Light Source (Oxfordshire, UK). Data collection took place at 100K, at a fixed wavelength of 0.9795 Å,

using a rapid read-out Pilatus 6M detector to record diffraction images. A total of 1800 images were recorded, over a rotation range of 180°, with a 0.1° rotation angle per image.

The data statistics are listed in **(Table 5.3)** and the diffraction images taken are shown in **(Figure 5.19B)**. The crystals diffracted to 2.3 Å resolution. They have a unit cell dimension of a = 121.56 Å, b = 121.56 Å, c = 69.83 Å, and a space group of $P3_121$. Cphy_1186 crystallises with 3 monomers per asymmetric unit, giving a Vm of 1.9 and a solvent content of 36 %.

Table 5.3: Crystallographic data statistics for native Cphy_1186.

Type	**Merged MTZ**					
Space group	$P3_121$					
Cell	121.56	121.56	69.83	90	90	120
Low resolution	105.2					
High resolution	2.3					
Number of lattices	1					
Number of reflections	18731					
Number of datasets	1					
Wavelength	0.9797					

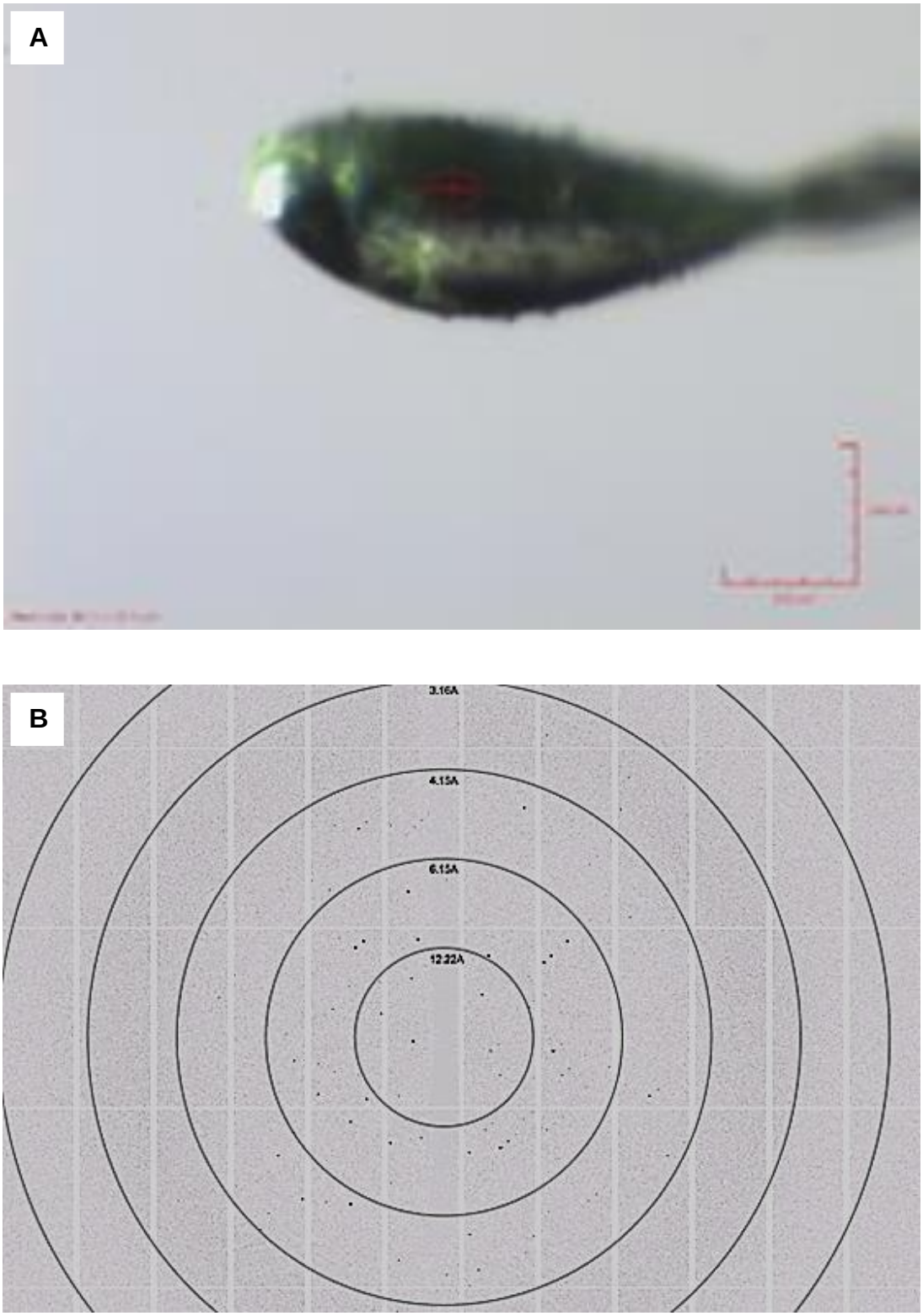


Figure 5.19: Cphy_1186 crystal shooting. A Single pentagonal Cphy_1186 crystal in a cryoprotected loop, about to be illuminated with an X-ray beam. The beam size to be used is 75.0 µm x 71.9 µm. B Diffraction image of a native Cphy_1186 crystal. The maximum resolution is 2.3 Å. The image was recorded at beamline I04-1, Diamond Light Source, Oxfordshire, UK.

5.2.3.4 Molecular replacement and subunit analysis

The protein sequence of *C. phytofermentans* Cphy_1186 and *C. freundii* PduT were aligned using *Clustal Omega*. The trimeric structure of Cphy_1186 was generated using the *PISA* software (CCP4 Suite). The structures were visualized and aligned by *PyMOL*.

The AutoMR and AutoBuild wizard from PHENIX produced a Cphy_1186 model using PduT (PDB ID: 3PAC) as the template. AutoBuild produced a model of 179 residues in six fragments with 5 waters, giving an R factor of 22.3 %, an R_{free} of 24.5 % and a correlation coefficient of 0.80. This model was used in cycles of refinement and rebuilding to yield the final Cphy_1186 structure at 2.3 Å resolution. The final model of Cphy_1186 has a clearly defined polypeptide backbone in the electron density map for 179 residues, and **Figure 5.20** shows an example of the electron density. The dark reddish colour of the purified Cphy_1186 sample suggests that the protein has retained its Fe–S iron–sulfur cluster during purification. Although the purification and crystallisation procedures were made in the absence of oxygen and that the crystals retained some coloured, the 4Fe-4S cluster is not seen in the final structure. That could be explained either by the loss of the iron–sulfur cluster during crystallisation or that the redox centre is very sensitive synchrotron radiation.

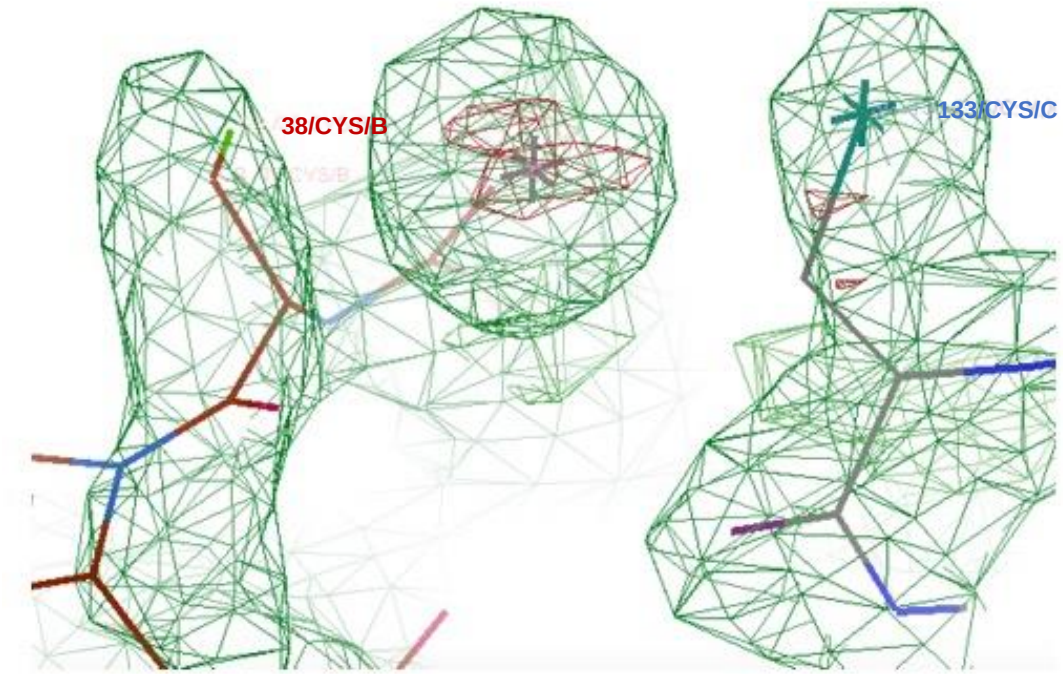
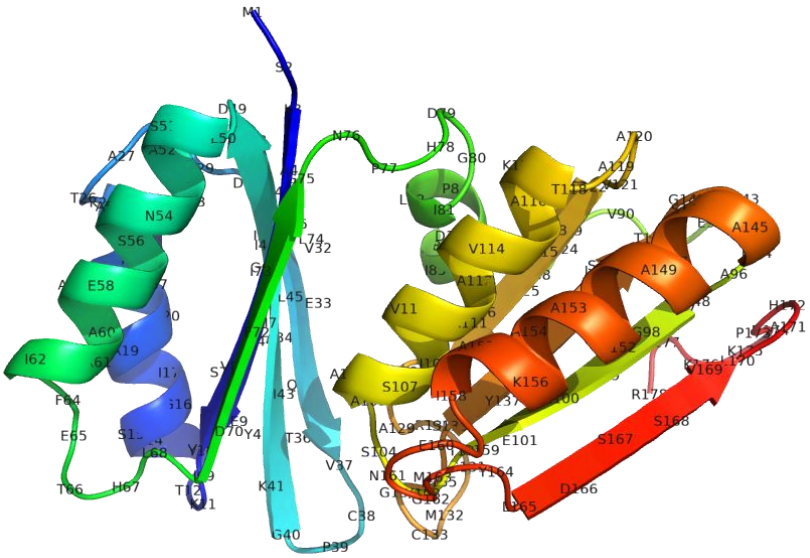


Figure 5.20: Electron density fitting of Cphy_1186 structure showing the region of the protein that incorporates Cys_{38} and Cys_{133}. This Figure was produced using PyMOL (DeLano and Lam, 2005).

The majority of shell proteins characterised as containing a single canonical BMC domain are approximately 90 residues in length. Cphy_1186 composed of 182 residues and contains two canonical BMC domains per subunit. Each domain of Cphy_1186 subunit consists of two β–α–β motifs, which are connected with the other domain by a beta hairpin forming an anti-parallel β-sheet **(Figure 5.21)**. The two-BMC-domains of the Cphy_1186 subunit is joined by a short α-helix and a β-turn. **Figure 5.21** also shows an alignment of quaternary structures of Cphy_1186 and PduT (PDB ID: 3PAC), where the general subunit appearance of Cphy_1186 aligns well. Cphy_1186 and PduT (PDB ID: 3PAC), which are composed of 182 and 184 residues respectively, share 67 % sequence similarity and structurally superimpose with an RMSD of 1.655 Å with 156 aligned residues.



```
        10         20         30         40         50         60
MSKAIGMVEY KTVSSGIMAA DLMVKTADID IVEAQTVCPG KYIILITGDL SAVNASVEAA

        70         80         90        100        110        120
RIQFETHLID SFILGNPHDG ILPAIYGASV VEEIEALGVL ETYSAASIIV AADVAAKTAA

       130        140        150        160        170        180
VELVEVRVAR GMCGKSYLML TGEIASVTAS IEAAKKAIGE NGMYLDSSVL AHPDKKLRNK

LM
```

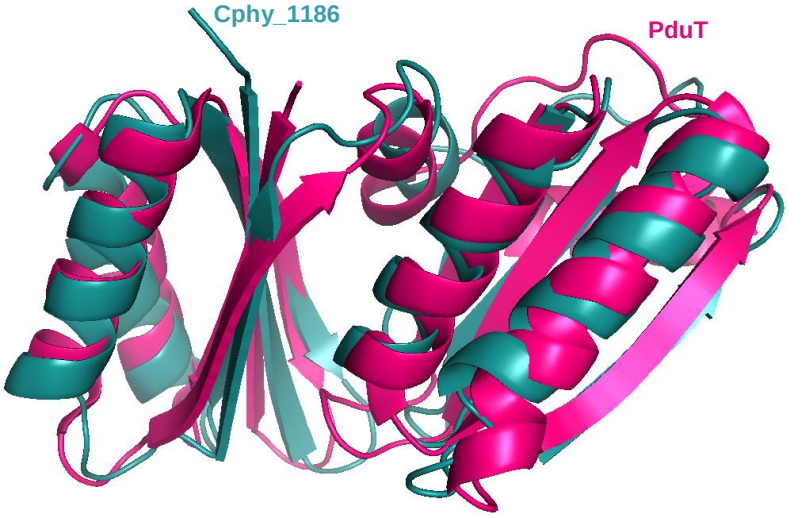


Figure 5.21: The tertiary structure of the Cphy_1186 subunit. On the top a cartoon representation of the tertiary structure of Cphy_1186 subunit which comprises two BMC repeats including the amino acid sequence of the Cphy_1186. On the bottom is the Cphy_1186 subunit and PduT subunit superimposed. This Figure was produced using PyMOL (DeLano and Lam, 2005).

5.2.3.5 Trimeric structure

Both the *C. freundii* PduT and *C. phytofermentans* Cphy_1186 contain a duplication of the canonical BMC domain. The Cphy_1186 trimer forms a flat, almost hexagonally shaped disk with a large central pore. One face of the Cphy_1186 trimer is concave and the other convex **(Figure 5.22)**. In addition, the pore loops at positions ~ 128 - 138 have very poor density and must be disordered in the crystal. The crystal structure analysis of Cphy_1186 revealed a tetrahedral geometry metal binding site at the interface of the trimers (each monomer binds a metal ion), which is most likely zinc (compute_mass: mass = 65.3900 u) and is coordinated by residues His-172 and Glu-95. The structure also contains one Na **(Figure 5.23)**.

The Cphy_1186 shell protein has a relatively large triangular-shaped central pore (~ 9 Å) that is lined with numerous cysteine residues. The cysteine residues most likely coordinate a 4Fe-4S cluster, leading to a suggested role for this cluster in an electron-transfer process **(Figure 5.24)**. The distances between cysteine residues are measured as Cys-38/Cys-133 ~ 5 Å per one subunit, and Cys-38/Cys-38 ~ 6 Å per two subunits **(Figure 5.24)**.

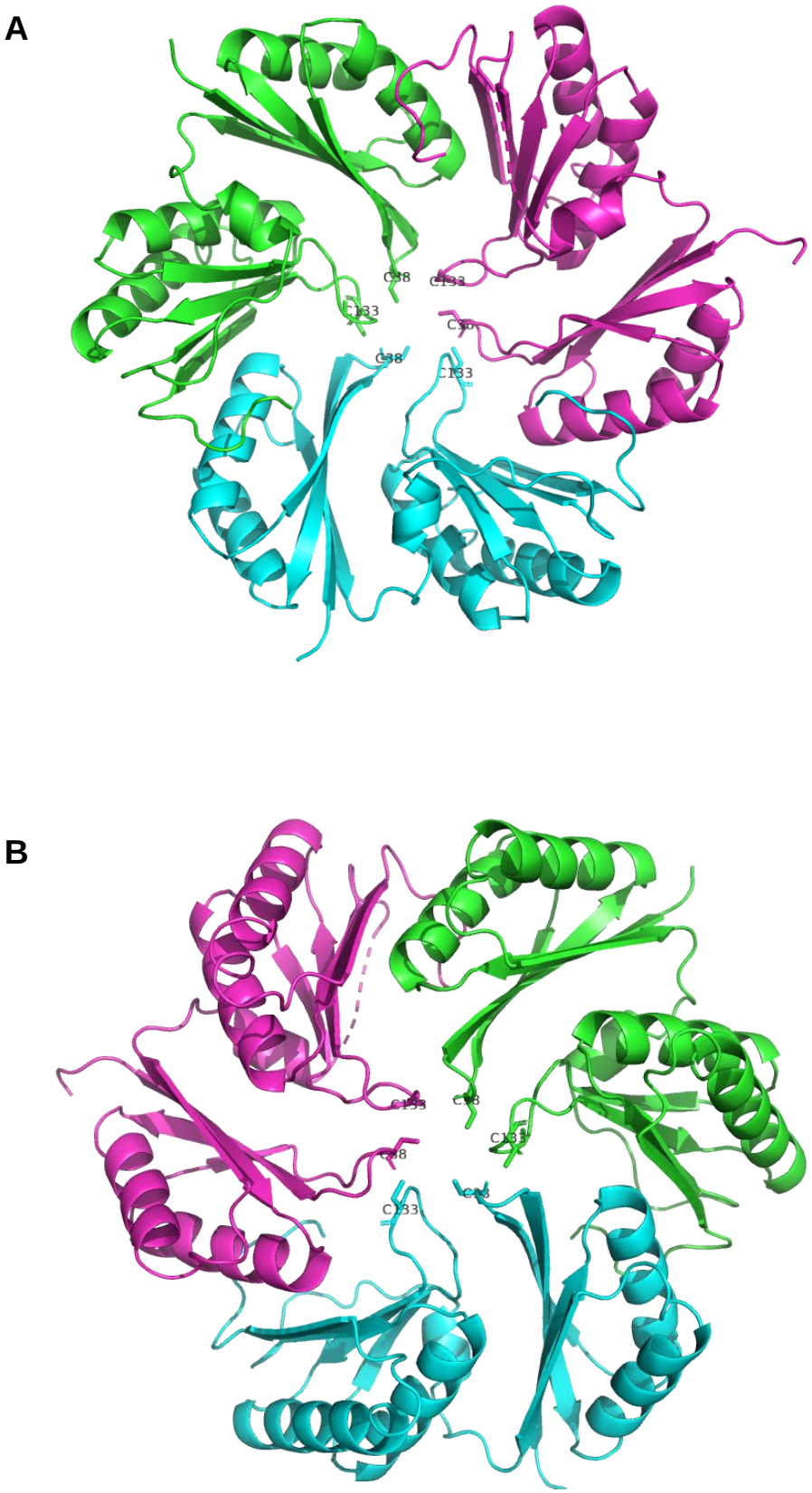


Figure 5.22: A quaternary and higher order structures involve in Cphy_1186. (A, B) The concave, side view and convex side of Cphy_1186. This Figure was produced using PyMOL (DeLano and Lam, 2005).

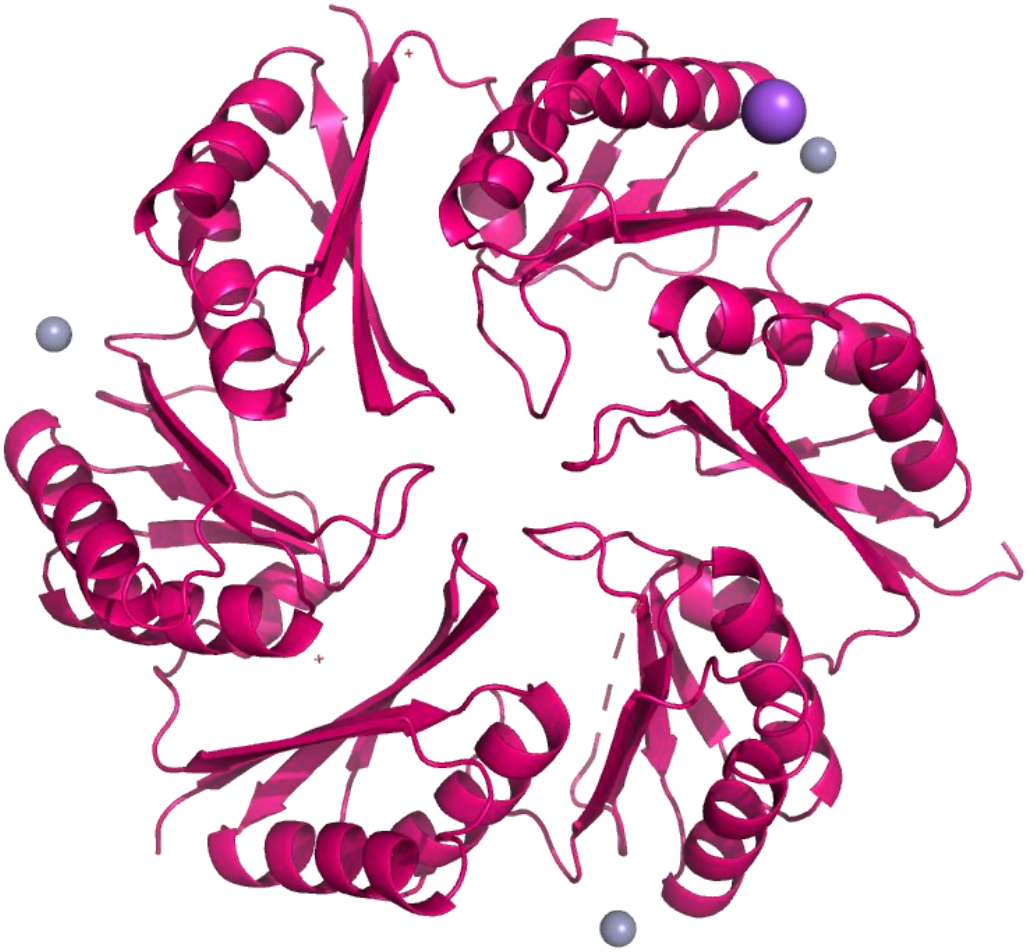


Figure 5.23: A quaternary structure of Cphy_1186 contains three atoms of Zn and one atom of Na. This side is the convex side of the trimer. This Figure was produced using PyMOL (DeLano and Lam, 2005).

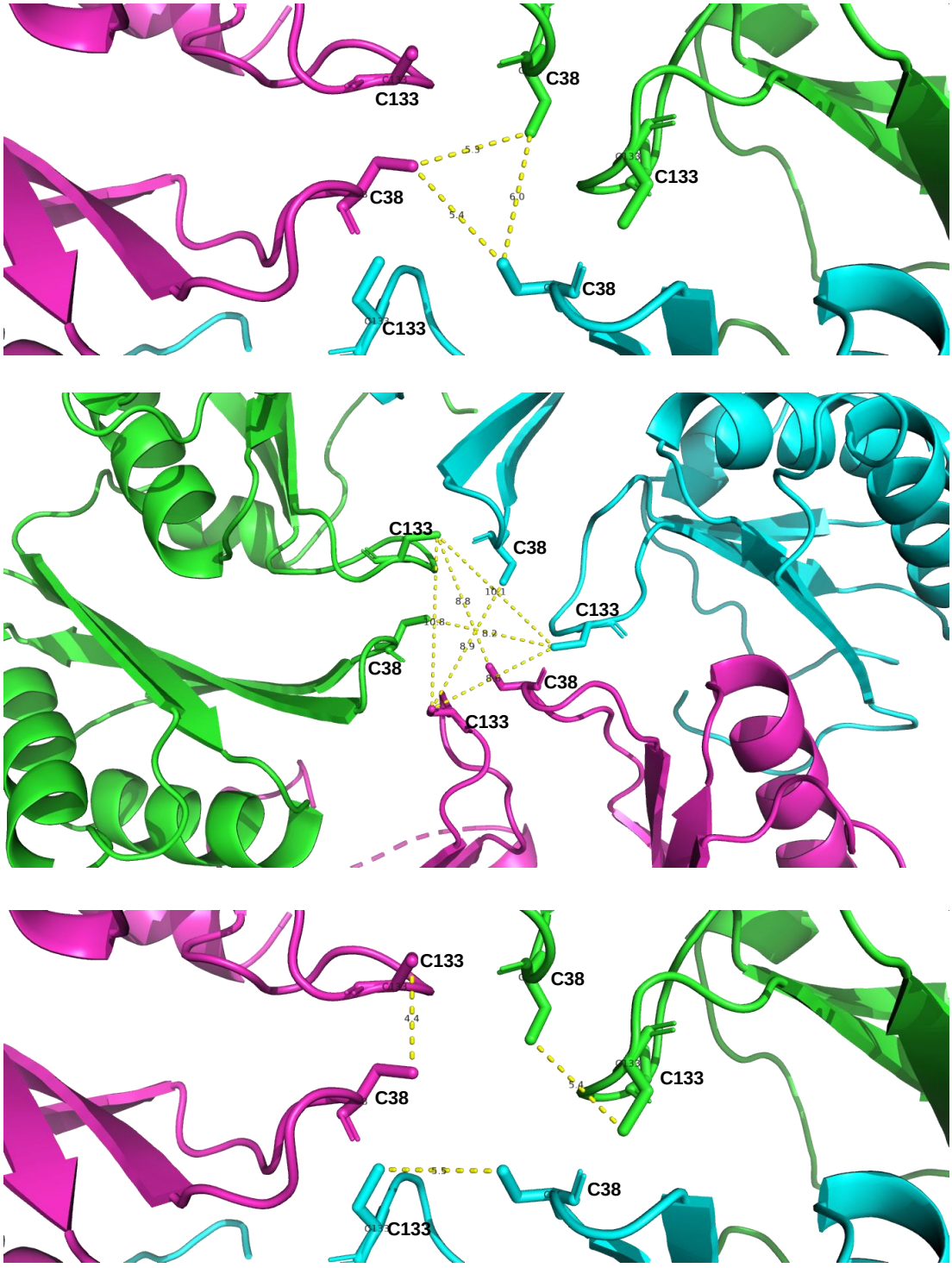


Figure 5.24: The subunit pore measurements of Cphy_1186 showing the positions of Cysteine residues that line the pore. This figure was produced using PyMOL (DeLano and Lam, 2005).

5.2.3.6 Molecular tiling of Cphy_1186

The crystal structure of *S. enterica* PduT showed them to form tightly packed molecular sheets with a trimer center-to-center spacing of 67.7 Å (Yeates *et al.*, 2010). By way of contrast, the crystal structure of Cphy_1186 reveals a higher order level of organisation with more spaced packing of trimers into nonuniform molecular sheets with a trimer centre-to-centre spacing of 71.9 Å **(Figure 5.25)**. This figure shows how the trimers of Cphy_1186 are packed in different levels and orientations. Each double trimer is joined together by the convex faces into a dihedral unit with 4.0 Å contacts between the conserved anti-parallel lysines and tyrosines **(Figure 5.26A).** According to the crystal structure, these units are joined in two ways: One by the zinc atoms, which are coordinated by residues His-172 and Glu-95, as shown in **Figure 5.26B**. In addition, the interactions between the trimer (convex)-trimer (concave) of the double trimer units are shown in **Figure 5.26C**, where the residue His-0 (from the convex side) forms 4.0 Å contacts with residues Phe-72, Ser-71 and Ile-73 (from the concave side). Moreover, Met-1 and Ser-51 (from the convex side) form 4.0 Å contacts with Asn-54 (from the convex side), whilst Asn-54 forms 4.0 Å contacts with Leu-50 (from the concave side) **(Figure 5.26C)**.

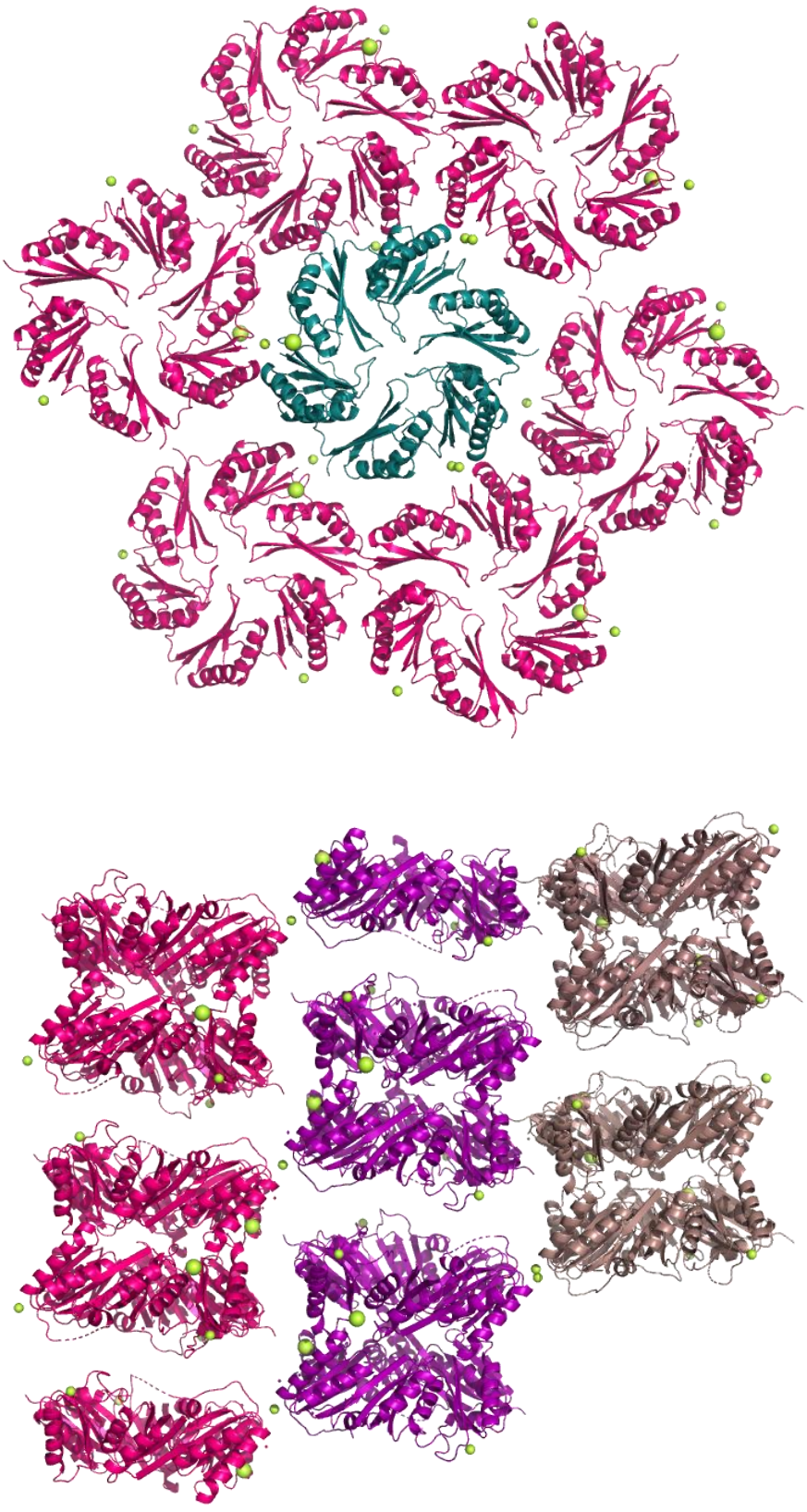


Figure 5.25: The crystal structure packing of Cphy_1186. In the top is the face view of the higher order packing of trimers into nonuniform molecular sheets. In the bottom is the side view of the crystal packing showing the way of trimer-trimer units packing. The small yellow atoms are Zn and the large yellow atoms are Na. This figure was produced using PyMOL (DeLano and Lam, 2005).

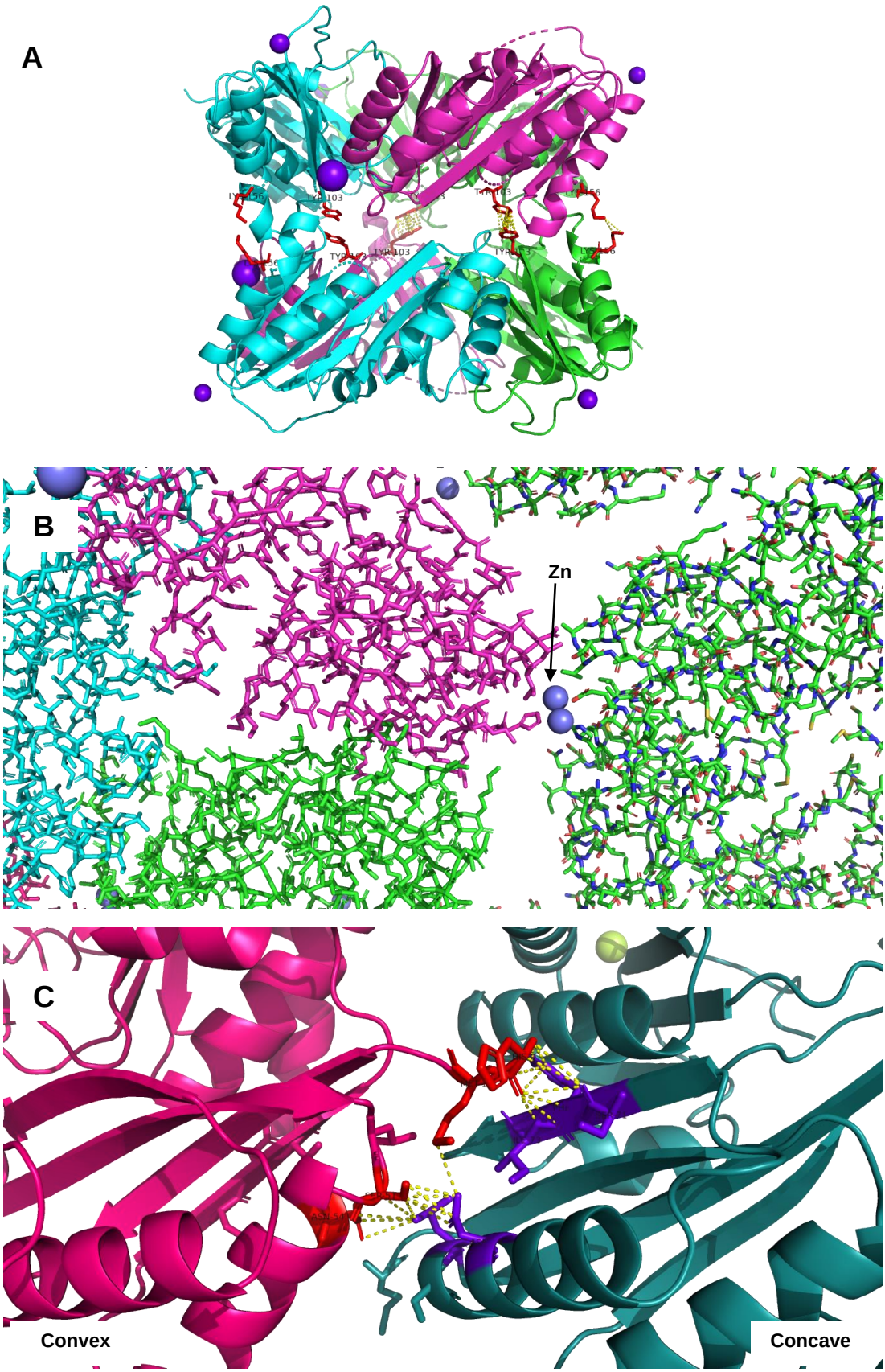


Figure 5.26: The crystal structure packing of Cphy_1186. A, a double trimer packing unit showing trimer-trimer interaction. B, subunit coloured representation of trimer-trimer interaction surface showing how zinc atoms drive the edges of the two trimers. C, trimer-trimer interaction surface showing the contacts between the edges of the two trimers.

5.2.3.7 Iron sulfur [4Fe-4S] binding site

Six structure of six trimeric shell proteins are now known including Cphy_1186 (the other being CsoS1D, EutL, EtuB, PduB, and PduT). The central pore of PduT exhibits a large triangular-shape opening. In addition, this shell protein has a binding site for a 4Fe-4S cluster in the central pore, where the β-hairpin loop between β2 and β3 point inwards to create a three-fold arrangement of C_{38} about the molecular three-fold axis (Pang *et al*., 2011). The Cphy_1186 shell protein also contains a large triangular-shape opening in the centre of the trimer. As will be discussed shortly, EPR and mutagenesis have been used to show that the Cphy_1186 shell protein has a binding site for 4Fe-4S cluster in the central pore. However, the Cphy_1186 trimer differs from PduT in that the β-hairpin loops from β2 to β3 as well as from β6 to β7 both point towards the pore, producing a six-fold arrangement of cysteine residues, C_{38} and C_{133}, about the molecular three-fold axis, which means it is possible that cysteine 133 may also participate in 4Fe-4S formation **(Figure5.27)**. There was no density for an Fe-S cluster in the pore but Cys 38 of PduT aligns perfectly with Cys 38 of Cphy_1186. The an-axis iron atom could be either up or down and is potentially available to bind another protein; the cluster is accessible from both sides and is therefore in a suitable location for single electron transfer across the shell of the bacterial microcompartment.

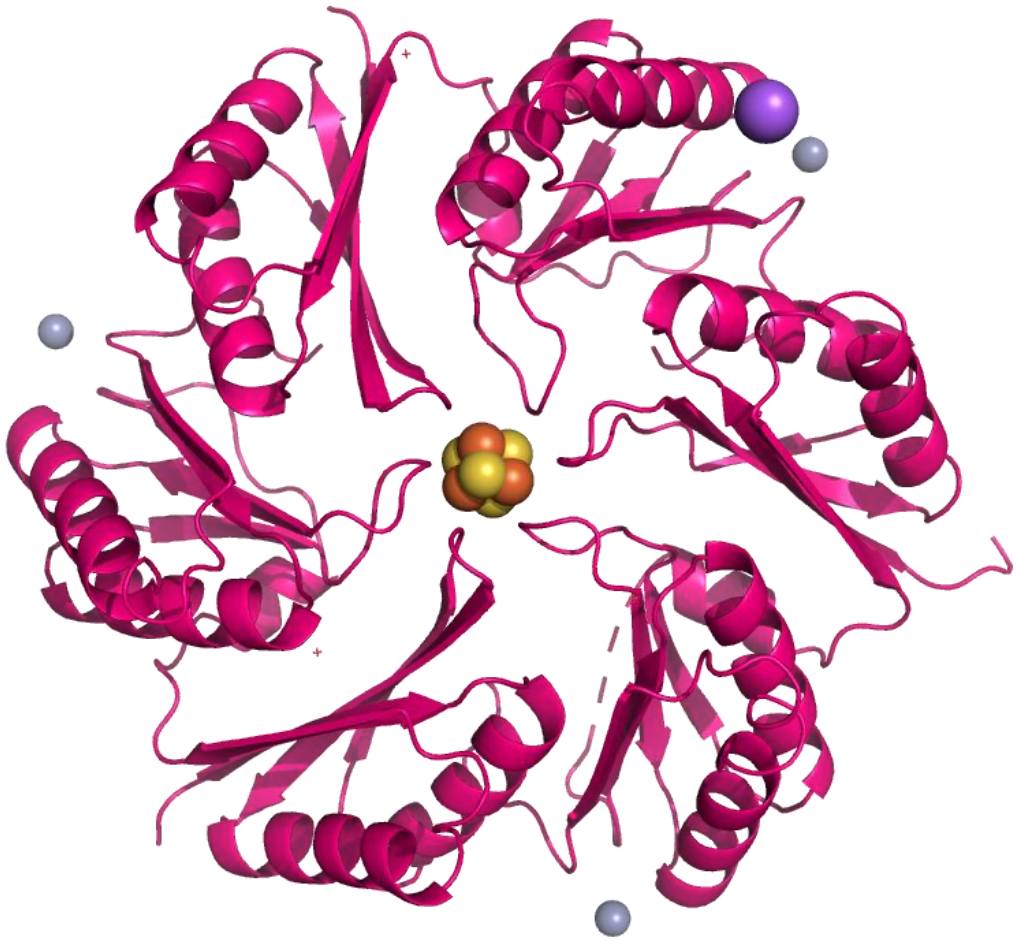


Figure 5.27: *Cphy_1186* and the binding site for 4Fe-4S. *Cphy_1186* trimer, showing the six β-hairpin loop that coordinate cysteines and pointing towards the central pore. This Figure was produced using PyMOL (DeLano and Lam, 2005).

5.2.3.8 Mutagenesis

To investigate further this possibility, two mutations were generated from the wild-type gene of Cphy_1186 to replace the cysteines residues (C_{38} and C_{133}) by conversion into alanine residues **(Figure 5.28)**. Thus, each mutation has just one cysteine residue instead of two in order to determine which of these cysteines is responsible for the brown colour and the Fe-S cluster. The mutant genetic variants of Cphy_1186 were ligated within pET14b plasmid (hexa histidine-Tag) and cloned into *E. coli* cells. The cells were expressed

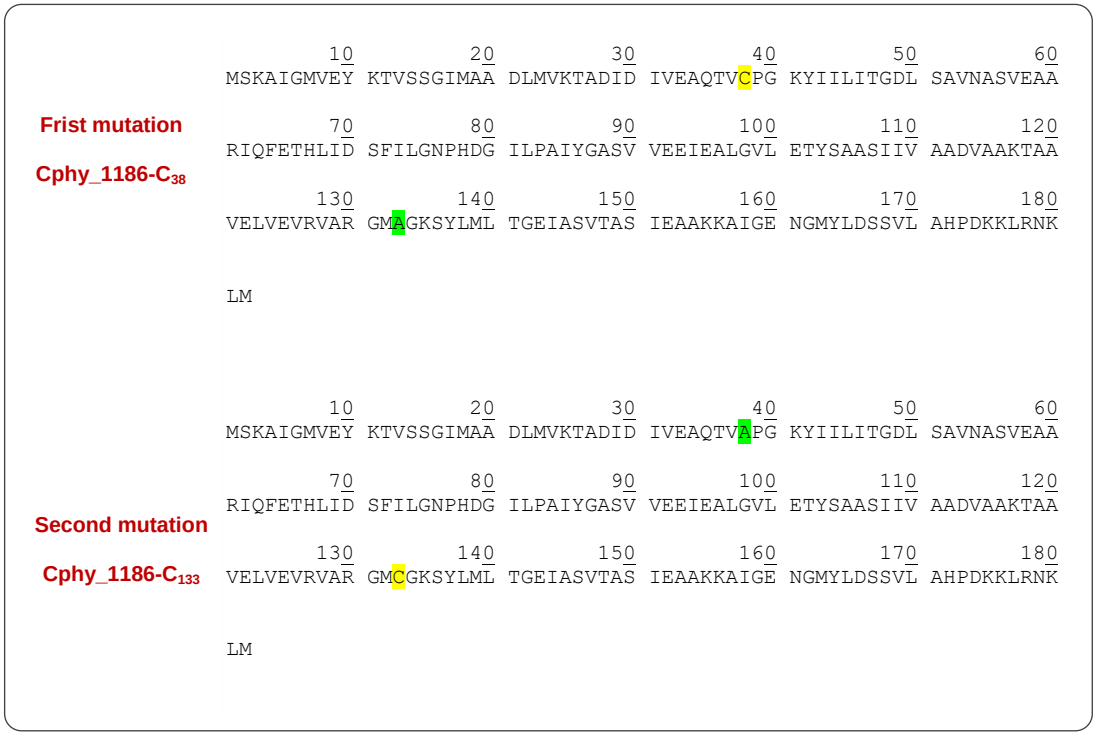


Figure 5.28: Mutagenesis of Cphy_1186 shell protein 182 amino acid sequence. The first mutation is to just replace C_{133} into Alanine. The second mutation is to just replace C_{38} into Alanine.

and induced overnight at 18 °C. For more details concerning the overproduction of protein see **Chapter 2**. Cultures were harvested and re-suspended with binding buffer **(Section 2.1.7.1)**. The cells were lysed aerobically by sonication and then centrifuged. Immediately, the supernatants were transferred into the Belle Technology glove-box under a nitrogen atmosphere and applied to separate columns filled with nickel charged resins (IMAC purification). The proteins were washed with two different imidazole concentrations (10 mM and 50 mM) buffers **(Section 2.1.7.1)**. After washing the columns to help remove unbound protein both variants displayed a distinct brown colour on the top layer. The brown proteins were isolated with a buffer containing 400 mM imidazole, followed

by PD-10 application. The collected fractions of both mutations were analysed by SDS-PAGE **(Figure 5.30)**, and were subject to UV-visible spectral analysis. The resulting spectra both had a broad absorption

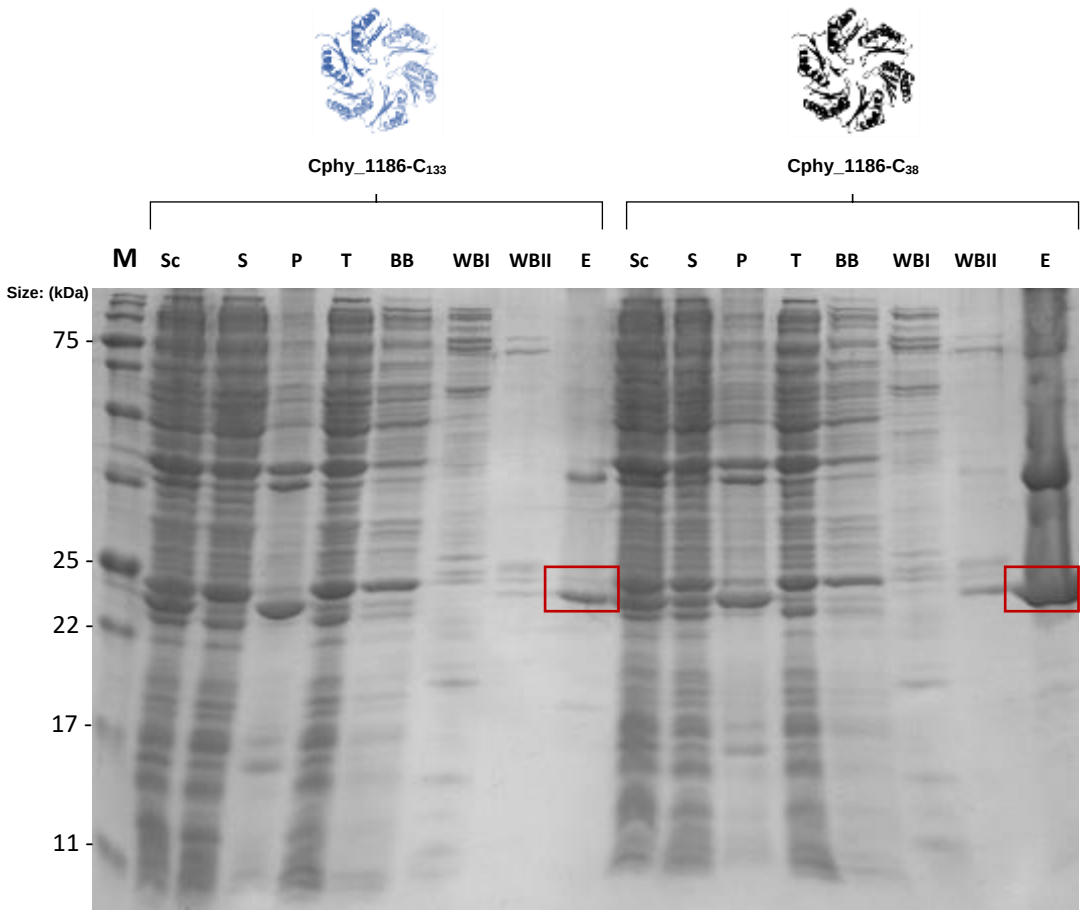


Figure 5.30: 15 % SDS PAGE for mutated Cphy_1186 IMAC fractions.

maximum at 420 nm **(Figure 5.31)**. The purified mutant variants were also subject to EPR analysis.

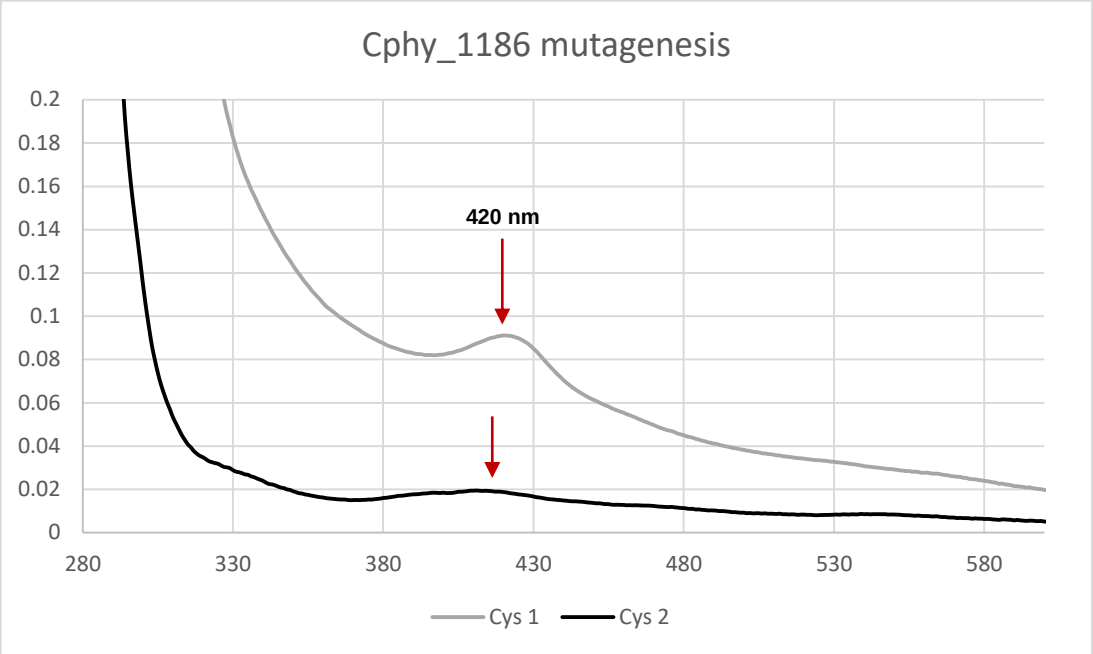


Figure 5.31: UV-visible absorbance spectrum of two Cphy_1186 mutations that purified anaerobically. The UV-visible spectrum of purified Cphy_1186 mutations in 20 mM Tris-HCl (pH 8.0), with 100 mM NaCl. Grey line represents the purified mutation of Cphy_1186-C_{38}. Black line represents the purified mutation of Cphy_1186-C_{133}.

5.2.3.9 Electron Paramagnetic Resonance (EPR)

To investigate about the Fe-S cluster further, EPR spectroscopy was undertaken on the two Cys to Ala mutants as well as for the wild-type Cphy_1186 protein. Whole cells containing the overproduced Cphy_1186 protein or its mutant variants were also subject to this analysis in order to distinguish between the cubic cluster of the trimer and the *E. coli* [2Fe-2S] ferredoxin **(Table 5.4)**.

A total of 14 samples of anaerobically purified Cphy_1186, Cphy_1186-C_{38}, Cphy_1186-C_{133}, whole cells, together with their reduced controls were analysed by EPR. Interestingly, purified Cphy_1186 exhibits the same [4Fe-4S]$^{1+}$ as the same sample that had been reduced with dithionite. The same

spectrum of the purified protein was also observed in whole cells **(Figure 5.32B, G)**.

Table 5.4: EPR samples of Cphy_1186 Fe-S cluster analysis.

Sample name	Details	Absorbance	Protein concentration
Cphy_1186	Anaerobically purified (have NOT oxidised any of them!)	at 280 nm = 8.4	20 mg/mL, 1 mM
Cphy_1186 R	Anaerobically purified/ reduced with 2 mM dithionite	-	20 mg/mL, 1 mM
WC Cphy_1186	Anaerobically purified/ whole cells (BL21(DE3) pLysS) expressing Cphy_1186	OD_{600} = 100	-
WC Cphy_1186 R	Anaerobically purified/ whole cells (BL21(DE3) pLysS) expressing PduT / reduced with 2 mM dithionite	OD_{600} = 100	-
Cphy_1186-C_{38}	Anaerobically purified/ mutated Cphy_1186 including one Cysteine (residue 38)	-	20 mg/mL, 1 mM
Cphy_1186-C_{38} R	Anaerobically purified/ mutated Cphy_1186 including one Cysteine (residue 38) / reduced with 2 mM dithionite	-	20 mg/mL, 1 mM
WC Cphy_1186-C_{38}	Whole cells (BL21(DE3) pLysS) expressing mutated Cphy_1186 including one Cysteine (residue 38)	OD_{600} = 100	-
WC Cphy_1186-C_{38} R	Whole cells (BL21(DE3) pLysS) expressing mutated Cphy_1186 including one Cysteine (residue 38)/ reduced with 2 mM dithionite	OD_{600} = 100	-
Cphy_1186-C_{133}	Anaerobically purified/ mutated Cphy_1186 including one Cysteine (residue 133)	-	20 mg/mL, 1 mM
Cphy_1186-C_{133} R	Anaerobically purified/ Cphy_1186 including one Cysteine (residue 133)/ reduced with 2 mM dithionite	-	20 mg/mL, 1 mM
WC Cphy_1186-C_{133}	Whole cells (BL21(DE3) pLysS) expressing mutated Cphy_1186 including one Cysteine (residue 133)	OD_{600} = 100	-
WC Cphy_1186-C_{133} R	Whole cells (BL21(DE3) pLysS) expressing mutated Cphy_1186 including one Cysteine (residue 133)/ reduced with 2 mM dithionite	OD600 = 100	-
pET14b	Whole cells (BL21(DE3) pLysS) expressing empty pET plasmid (control)	OD_{600} = 100	-
pET14b R	Whole cells (BL21(DE3) pLysS) expressing empty pET plasmid/ reduced with 2 mM dithionite (control)	OD_{600} = 100	-

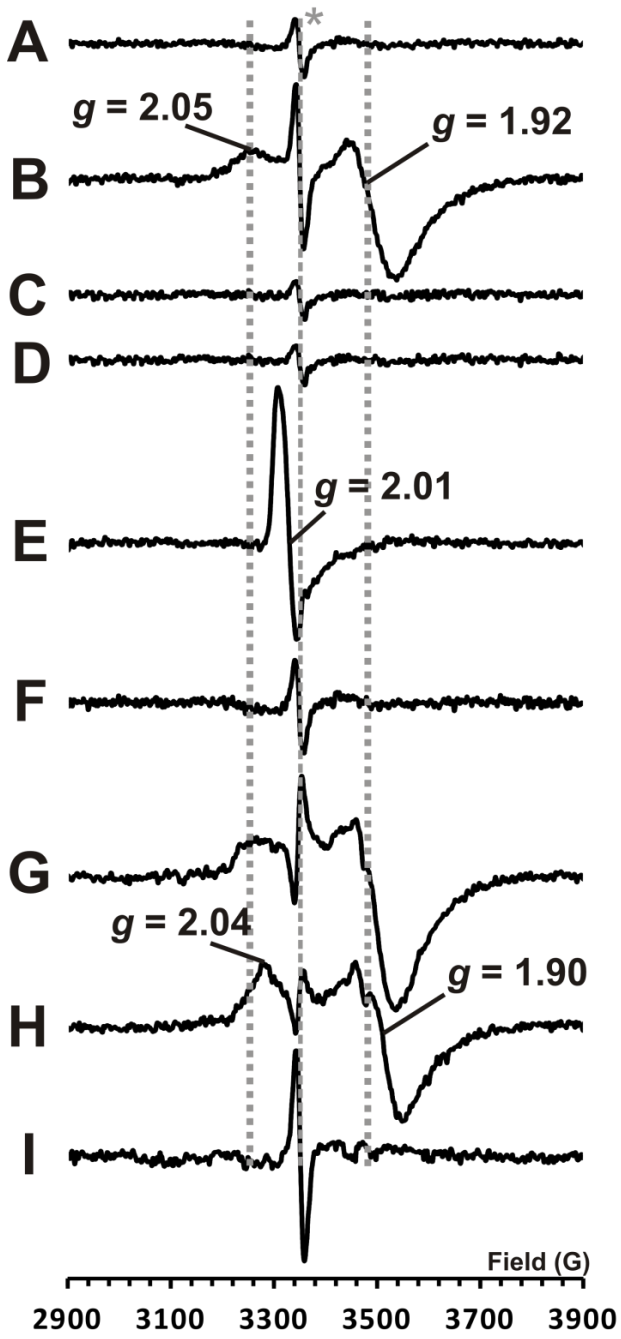


Figure 5.32: X-band CW EPR spectra of Cphy_1186 anaerobically purified (A-F) and expressed in whole cells (G-I) showing [4Fe-4S] signal: (A) Cphy_1186 as purified. (B) Cphy_1186 reduced with dithionite. (C) mutation Cphy_1186-C_{38} as purified. (D) mutation Cphy_1186-C_{38} as dithionite reduced. (E) mutation Cphy_1186-C_{133} as purified. (F) mutation Cphy_1186-C_{133} as dithionite reduced. (G) Cphy_1186 expressed in whole cells, dithionite reduced minus as purified. (H) mutation Cphy_1186-C_{38} expressed in whole cells, dithionite reduced minus as purified. (I) mutation Cphy_1186-C_{133} expressed in whole cells, dithionite reduced minus as purified. Experimental conditions: microwave power 1 mW, modulation amplitude 7 G, temperature 15 K, each spectrum is the sum of four scans.

However, the Cphy_1186-C_{38} (where Cys133 has been mutated) sample has a slightly modified $[4Fe\text{-}4S]^{1+}$ cluster with g = 2.04 in (H) that does not survive the purification process (no signal in D). With the Cphy_1186-C_{133} (Where Cys38 has been mutated), a $[3Fe\text{-}4S]^{1+}$ cluster is observed in vitro with the large signal at g = 2.01 in E, but this may be just an artefact of purification as no such signal is observed in whole cells in (I).

The reason why reduced minus oxidised spectra were used for the whole cell studies is because the purified Cphy_1186 do not exhibit quite the same transition metal EPR signal background as the whole cells expressing the protein. Thus, the reduced minus oxidised approach is more effective at revealing the presence of signals that can be attributed to Cphy_1186 than subtraction of the purified protein only background. In **Figure 5.33** some of the signals observed in the spectrum of Cphy_1186 expressed in whole cells

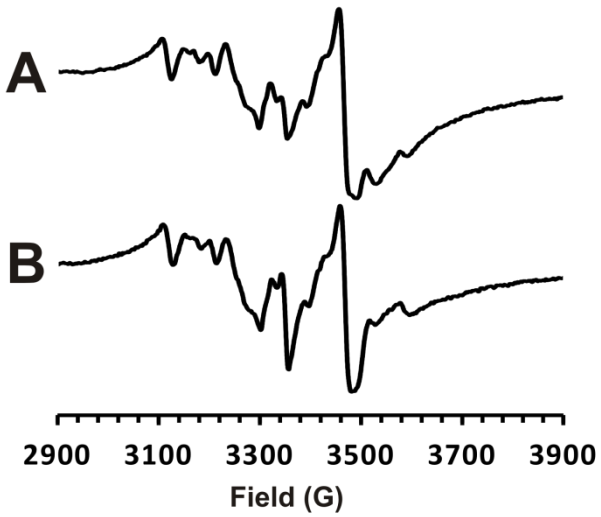


Figure 5.33: X-band CW EPR spectra of Cphy_1186 expressed in whole cells with and without reduction showing [4Fe-4S] signal: (A) is the Cphy_1186 in whole cells reduced with dithionite. (B) is the spectrum from the same cells without reduction.

arise from transition metal centres within the bacterial cytoplasm or periplasm, particularly Mn^{2+} and the *E. coli* [2Fe-2S] ferredoxin. All samples exhibit small amounts of radical signal at *g* = 2.00, marked with *, which appears reversed in (G and H) because the signal is larger in the as purified than in the reduced samples. This suggests that both cysteine residue from Cphy_1186 protein ligate a 4Fe-4S cluster, a highly unusual arrangement but consistent with the proposed homotetrameric structure.

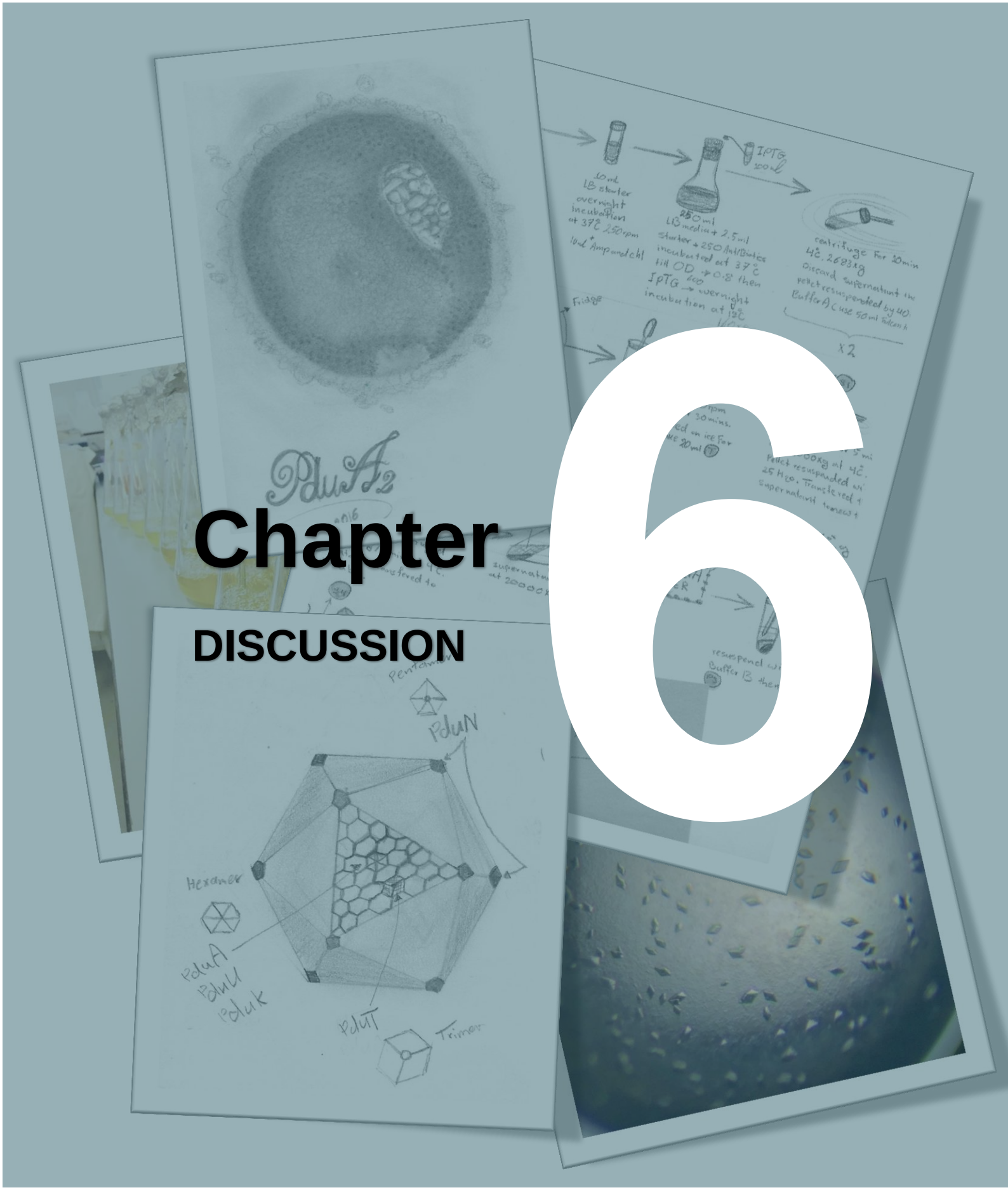
5.3 Conclusion

The crystal structure of Cphy_1186 shows a trimeric assembly of subunits each containing a double-BMC-domain with a large triangular shape central pore. The trimer and its pore have a similar size to PduT (PDB ID: 3PAC), and align well. Cphy_1186 has a central pore that is unique. It has six cysteine residues that they are positioned such that they are able to bind three of the four iron atoms of a 4Fe-4S cluster, leaving the fourth iron atom free for interaction with another protein such as Cphy_1185 (PduS-like). Initially, sequence analysis by Phyre[2] has indicated the presence of such a [4Fe-4S] on Cphy_1185, which has 58 % similarity to PduS. The PduS is thought to act as a corrin reductase for the formation of adenosylcobalamin (Petit *et al.*, 2013). It has been proposed that PduS employs its 4Fe-4S cluster to absorb unneeded electrons inside the microcompartment and to pass them out of the shell through PduT (Parsons *et al.*, 2008).

Experimentally, this electron transfer could not be observed between PduS and PduT. Many attempts were made to form the crystal structure of

Cphy_1186 anaerobically with intact Fe-S cluster, and the attempts failed to generate crystals with electron density for the Fe-S cluster.

Finally, it will be interesting to see if Cphy_1186 interacts with other enzymes such as Cphy_1185 (PduS-like) by purifying a tagged enzyme (Cphy_1185) and looking for an interaction with an untagged shell protein (Cphy_1186) with IMAC. If the purification of the interaction succeeds, the next step would be co-crystallization of the two purified proteins in order to provide experimental evidence about the role of the Fe-S centre in of Cphy_1186.



Chapter 6

DISCUSSION

6.1 Discussion

The discovery of bacterial microcompartments (BMCs) in 1950s revolutionised our ideas on the cell biology of prokaryotic cells. The presence of large icosahedral complexes within the cytoplasm indicated that some bacteria had the ability to form proteinaceous compartments, thereby demonstrating that organelles are not just a defining characteristic of eukaryotic cells (Shively *et al.*, 1973; Chun *et al.*, 2014). These initial BMCs were shown to be carboxysomes, harbouring RuBisCO and carbonic anhydrase to assist with carbon fixation. Several decades later, similar proteinaceous structures were observed in *S. enterica* that were associated with the utilisation of compounds such as propanediol and ethanolamine. These BMCs were termed metabolosomes and the evidence for the presence of propanediol utilisation (Pdu) and ethanolamine utilisation (Eut) BMCs throughout many bacterial species is provided by the many bacterial genomes that have been sequenced (Bobik *et al.*, 1999). In 2013 a further BMC-encoding locus many genes homologous to the *S. enterica Pdu* BMC system, was reported with studies undertaken in *Clostridium Phytofermentans.* This locus in thought to be involved in fucose/rhamnose metabolism and when *C. phytofermentans* was grown on these sugars TEM analysis of thin sectioned cells revealed polyhedral structures, consistent with the presence of BMCs (Petit *et al.*, 2013). The 13 genes within this operon are given the acronym *Cphy*.

This study contributes to the growing wealth of knowledge concerning the structural proteins that help form the proteinaceous outer casing of the BMC,

the shell proteins. Six of the genes within the operon encode for shell proteins. Initially, the six shell proteins were cloned individually and then collectively to allow the production of their encoded proteins. Five of the *C. phytofermentans* six shell proteins of were found to be soluble and were purified successfully, the exception being Cphy_1176, which was produced as a largely insoluble protein. ***Phyre²*** sequence analysis of the Cphy_1176 shell protein showed that it displayed 57 % similarity to the *S. enterica* PduU shell protein (PDB ID: 3CGI) and indicated that the protein subunit contains an unusual stranded barrel, which is composed of short α-helixes instead of β-sheets **(Figure 3.6)**. This strongly suggests that Cphy_1176 is a hexamer formed from six single-BMC-domain subunits and can be termed a BMC-H shell protein.

Characterisation of some of the other shell proteins, including Cphy_1180, Cphy_1181, and Cphy_1182 indicated that they too formed hexamers and were therefore likely to be BMC-H proteins. Cphy_1186 was found to form a trimer and is therefore a BMC-T protein. Finally, Cphy_1184 was observed to form pentamers meaning that it belonged to the BMC-P class of shell protein. Subsequent protein crystallography and X-ray structure analysis of Cphy_1182 and Cphy_1186 confirmed that these shell proteins are belong to BMC-H and BMC-T classes, respectively.

TEM analysis has revealed many different types of structures formed in the cytoplasm of *E. coli* cells from overproduction of individual shell proteins. These structures indicate the inherent ability of the shell proteins to form large macromolecular structures. These structures also provide some information about the possible role of the individual shell proteins. For

example, overproduction of Cphy_1176 results in the appearance of filamentous structures, similar to the filaments seen with PduA overproduction. Such filaments are generated through the ability of the hexameric tiles to link together in a curved fashion such that a tube is formed **(Figure 4.2)**.

Some of the *C. phytofermentans* shell proteins when overproduced form defined structures including rolled sheets. This is observed with the overproduction of Cphy_1181, which forms elongated sheets. These rolled sheets arise because of the specific. The side-by-side interaction allow for a planar elongation of the structure, whilst the end-to-end hexamer interaction tilts slightly to form a roll of approximately 200 nm in diameter.

Overproduction of Cphy_1182 also results in the formation of long filaments that are more defined than those seen with Cphy_1176, appearing more as nanotubes. Such structure would be consistent with Cphy_1182 forming the edges that join the facets of the metabolosome. The structure of Cphy_1182 reveals that the protein forms a dimer in the crystal lattice with zinc atoms located in between the subunits. This was unexpected as most of these BMC-H proteins tend to crystallise as hexamers. It would appear that the zinc atoms may drive the formation of the dimer. Under physiological conditions, when the protein is extracted analysis by gel filtration chromatography indicates that the protein is a hexamer. The crystal structure of Cphy_1182 also highlights four residues found at the edge of the subunit, Arg-157, Pro-158, Asp-128 and Asn-107, which may play an important role in the hexamer-hexamer interaction. It would appear that these interactions of these residues are significant in both tiling (angle of 0°) and nanotubes

(angle of 25° or 30°) formation. Indeed, nanotubes formed from BMC-domain protein open up many exciting possible applications, such as alternatives to carbon nanotubes, which are applied in many scientific areas of nanotechnology, electronics and optics.

Cphy_1184 is the only structural subunit of *C. phytofermentans* to not form any structures when overproduced, and that may be because of the high solubility of the protein **(Figure 3.24)**. Cphy_1184 is thought to form the pentameric vertices of the metabolosome. Interestingly, the purified recombinant Cphy_1184 purified with a yellow colour, although EPR analysis did not show any signal for Fe-S centre. The pigment associated with Cphy_1184 therefore remains unknown but there has not been any previous report of colour associated with a BMC-P protein.

Purified Cphy_1186 had a brown colour and the UV-visible spectrum suggested the presence of an iron sulphur cluster [Fe-S] with broad peaks around 420 nm. An EPR analysis revealed the formation of a X-band signal that was consistent with the presence of a $[4Fe\text{-}4S]^{1+}$ centre, the protein contains two cysteine residues (C_{38} and C_{133}) and mutagenesis of these indicates that both are associated with the formation of the Fe-S centre. As expected, the crystal structure of Cphy_1186 reveals it to have trimeric arrangement of subunits each of which contains a double-BMC-domain. The pseudo-hexamer has a central pore where the two cysteine residues are located. The pore therefore is lined with six cysteine residues, more than enough than the four that are required for Fe-S centre formation. Although the anaerobically produced crystal was coloured, no electron density for the iron-sulphur cluster was found in the structure of Cphy_1186. Consequently,

a computational model of Cphy_1186 bound to a [4Fe-4S] cluster was prepared. Cphy_1186 may interact with other Fe-S proteins such as Cphy_1185 (PduS-like), which sequence analysis also indicates the presence of a [4Fe-4S]. In the *C. freundii* Pdu metabolosome it has been suggested that PduS uses its two [4Fe-4S] centre to absorb free electrons and pass them onto PduT (Pang *et al*., 2011). However, the *C. phytofermentans* BMC houses a glycyl-radical propanediol dehydratase (Cphy_1174) (Grp-type) (Petit *et al.*, 2013). Thus the Fe-S centre may also somehow be involved in activating the 1,2-propanediol dehydratase activase (Cphy_1175), which requires the presence of [4Fe-4S] (Shisler and Broderick, 2014).

In vivo overproduction of Cphy_1186 results in the formation of relatively long, thick, swirl bodies. Moreover, the crystal structure of Cphy_1186 reveals double stacking trimers similar to that seen in the intact shell of the *Haliangium ochraceum* (PDB: 5V76) (Sutter *et al.*, 2017). In this case the concave sides of the double stacking trimers (BMC-T) face each other to form an air-lock. However, a key difference is that in Cphy_1186 the convex sides of the double stacking trimers face each other. Sequence alignment between Cphy_1186 and the *H. ochraceum* BMC-T (PDB: 5V76) reveals 23 % identity and 38 % similarity. All this information suggests that Cphy_1186 assembles in the intact shell of the *C. phytofermentans* as a double stacking trimer, where the convex sides of the lower trimers face outwards. A summary of the proposed role and characteristics of the six shell proteins is highlighted in **Table 6.1**.

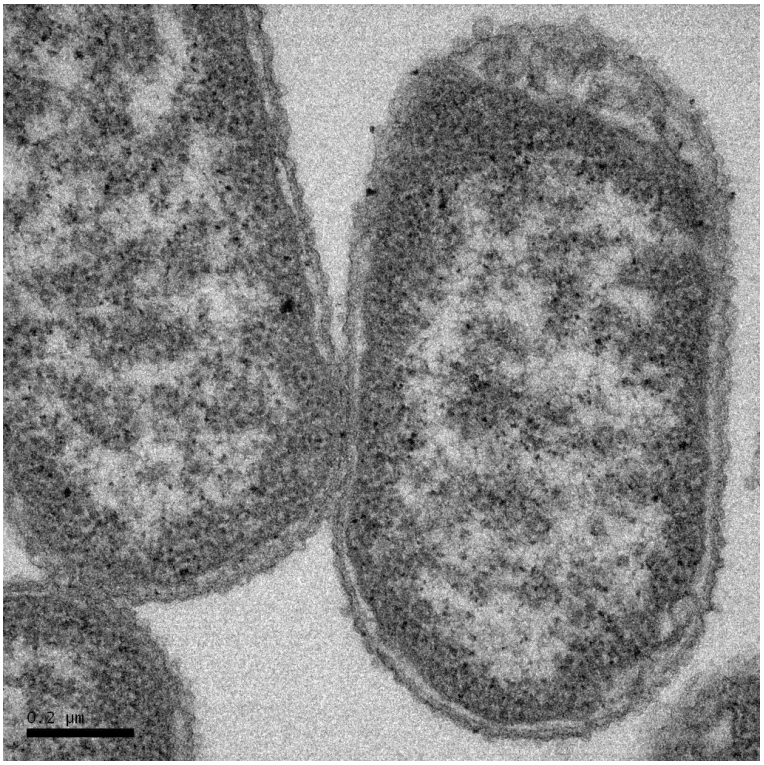
Table 6.1: Summary of the shell protein characteristics.

Protein	*S. enterica* (% similarity)	solubility	Colour	Expression	Possible role	Cofactors
Cphy_1176	PduU (57)	insoluble	Clear	Long filaments	Forms the flat facets of the BMC	-
Cphy_1180	PduA (51)	Partly soluble	Clear	Inclusion bodies	Forms the flat facets of the BMC	-
Cphy_1181	PduK (53)	Soluble	Clear	Large, long rolled sheets	Forms the flat facets of the BMC	-
Cphy_1182	PduA (90)	Soluble	Clear	Nanotubes around 25 nm in diameter	Forms the edges of the flat facets of the BMC	-
Cphy_1184	PduN (60)	Soluble	Yellow	None	Forms vertices of the BMC	-
Cphy_1186	PduT (66.3)	Soluble	Brown	Relatively long thick swirls	Substrate, product and electron transport	$[4Fe\text{-}4S]^{+1}$

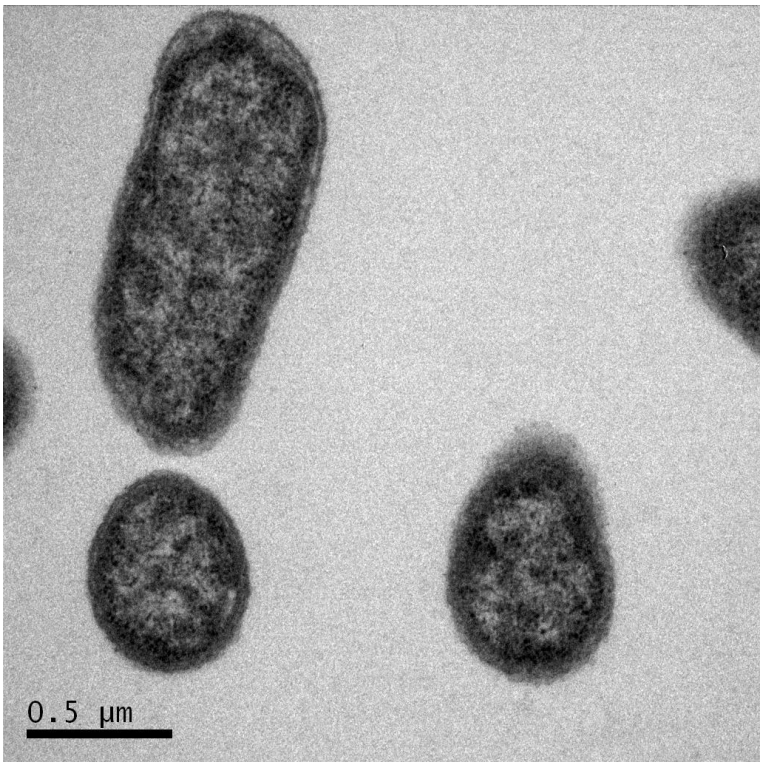
Bacterial cells form BMCs for a range of metabolic processes and in so doing protects the cell from toxic intermediates, increase the internal substrate concentration and prevent the release of volatile intermediates. In addition, the *Cphy* metabolosome may not only increase the metabolic efficiency of the enclosed enzymes, but may also protect the cell from radical intermediates associated with radical enzymatic reactions, and at the same time maintain an anaerobic environment for the Fe-S clusters to prevent them from oxidative damage.

The first attempt at engineering empty metabolosomes from the six shell proteins that are involved in the formation of the *C. phytofermentans* metabolosome has been presented. Although ultimately this was unsuccessful the associated research has given greater insight into the structure/function relationship between the individual components. BMCs hold considerable potential for use in many applications including nanomedicine, energy production and for increasing the efficiency of cells as cell factories for the production of commodity chemicals. There are opportunities whereby cellular nanotechnology can be used to help develop areas of biotechnology and thus BMCs have a significant future in biological research.

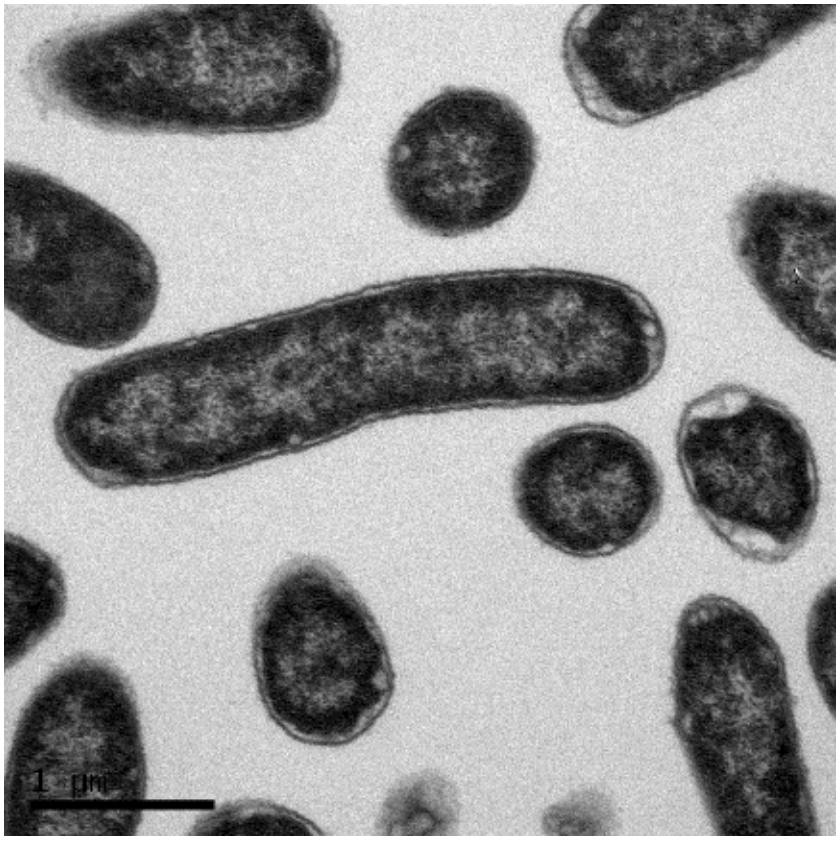
Images of controls



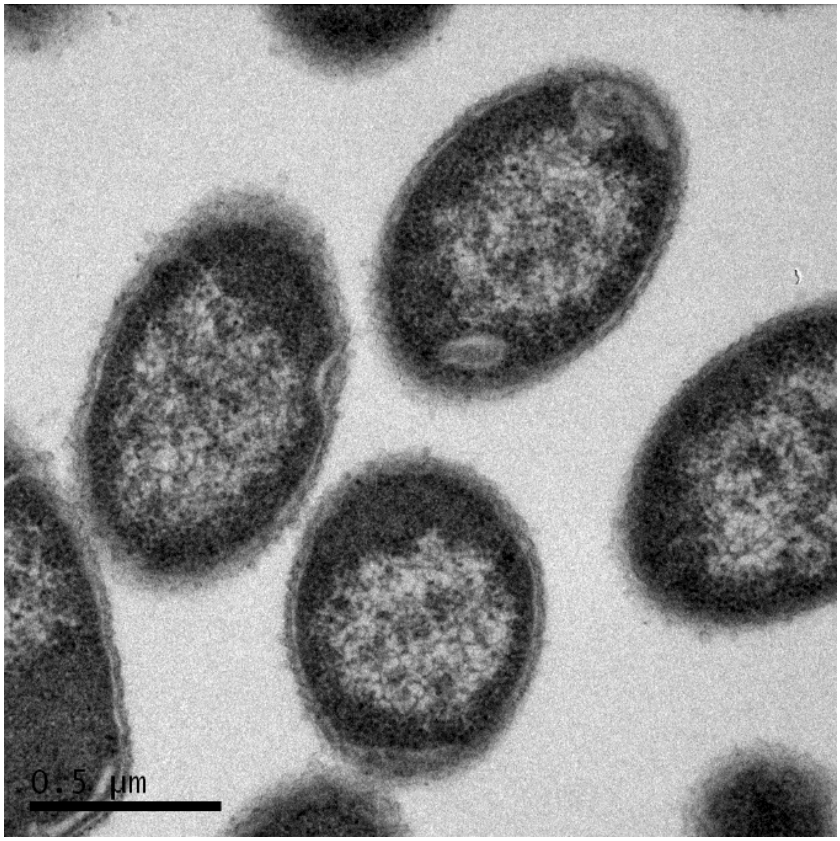
Control section of *E. coli* containing pET3a. The cells were induced overnight with 200 µM IPTG.



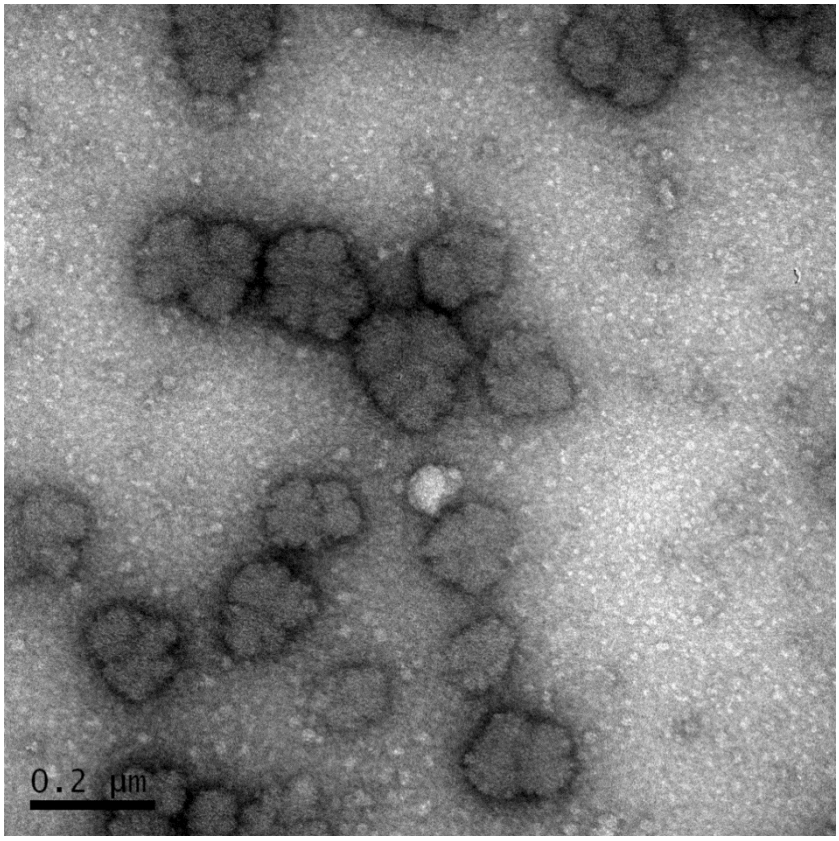
Control section of *E. coli* containing pLysS. The cells were induced overnight with 200 µM IPTG.



Control section of *E. coli* containing pET3a. The cells were induced for 2 hours with 400 µM IPTG.



Control section of *E. coli* containing pET3a. The cells were induced for 2 hours with 200 µM IPTG.



Control of Figure 4.14.

References

Aussignargues, C. *et al.* (2015) 'Bacterial microcompartment assembly: The key role of encapsulation peptides', *Communicative and Integrative Biology*, 8(3), p. 1039755.

Axen, S. D., Erbilgin, O. and Kerfeld, C. A. (2014) 'A Taxonomy of Bacterial Microcompartment Loci Constructed by a Novel Scoring Method', *PLoS Computational Biology*. Edited by M. M. Tanaka. Public Library of Science, 10(10), p. e1003898.

Badger, M. R. and Bek, E. J. (2008) 'Multiple Rubisco forms in proteobacteria: Their functional significance in relation to CO_2 acquisition by the CBB cycle', *Journal of Experimental Botany*, 59(7), pp. 1525–1541.

Beudeker, R.F., Cannon, G.C., Kuener, J.G., and Shively, J.M. (1980) 'Relations between d-ribulose-1,5-bisphosphate carboxylase, carboxysomes and CO_2 fixing capacity in the obligate chemolithotroph *Thiobacillus neapolitanus* grown under different limitations in the chemostat', *Archives of Microbiology*, 124, 185-189.

Bobik, T. A. *et al.* (1999) 'The propanediol utilization (pdu) operon of *Salmonella enterica* serovar Typhimurium LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme B_{12}-dependent 1,2-propanediol degradation', *Journal of Bacteriology*, 181(19), pp. 5967–5975.

Bobik, T. A., Ailion, M. and Roth, J. R. (1992) 'A single regulatory gene integrates control of vitamin B_{12} synthesis and propanediol degradation', *Journal of Bacteriology*, 174(7), pp. 2253–2266.

Bobik, T. A., Lehman, B. P. and Yeates, T. O. (2015) 'Bacterial microcompartments: Widespread prokaryotic organelles for isolation and optimization of metabolic pathways', *Molecular Microbiology*, 98(2), pp. 193–207.

Brinsmade, S. R., Paldon, T. and Escalante-semerena, J. C. (2005) 'Minimal Functions and Physiological Conditions Required for Growth of', *Microbiology*, 187(23), pp. 8039–8046.

Cai, F. *et al.* (2014) 'Engineering bacterial microcompartment shells: Chimeric shell proteins and chimeric carboxysome shells', *ACS Synthetic Biology*, 4(4), pp. 444–453.

Campbell, G. *et al.* (2006) 'Whole-Genome Transcription Profiling Reveals Genes Up-Regulated by Growth on Fucose in the Human Gut Bacterium "*Roseburia inulinivorans*"', *Journal of Bacteriology*, 188(12), pp. 4340–4349.

Cannon, G. C. *et al.* (2001) 'Microcompartments in Prokaryotes: Carboxysomes and Related Polyhedra', *Applied and Environmental Microbiology*, 67(12), pp. 5351–5361.

Cannon, G.C. and Shively, J.M. (1983) 'Characterization of a homogenous preparation of carboxysomes from *Thiobacillus neapolitanus*', *Archives of Microbiology*, 134, 52-59.

Chen, P., Andersson, D. I. and Rot, J. R. (1994) 'The Control Region of the pdu/cob Regulon in *Salmonella typhimurium*', *Journal of Bacteriology*, 176(17), pp. 5474–5482.

Cheng, S. *et al.* (2008) 'Bacterial microcompartments: Their properties and paradoxes', *BioEssays*, 30(11-12), pp. 1084–1095.

Cheng, S. *et al.* (2011) 'Genetic analysis of the protein shell of the microcompartments involved in coenzyme B_{12}-dependent 1,2-propanediol degradation by *Salmonella*', *Journal of Bacteriology*, 193(6), pp. 1385–1392.

Chun, S. *et al.* (2014) 'Diverse Bacterial Microcompartment Organelles', *Microbiology and Molecular Biology Reviews*, 78(3), pp. 438–468.

Crowley, C. S. *et al.* (2008) 'Structure of the PduU Shell Protein from the Pdu Microcompartment of Salmonella', *Structure*, 16(9), pp. 1324–1332.

DeLano, W.L. and Lam, J.W. (2004) 'PyMOL: A communications tool for computational models', *Abstracts of Papers of the American Chemical Society*, 230, U1371–U1372.

Ellis, R.J. (2010) 'Tackling unintelligent design', *Nature,* 463, 164-165.

Fejer, S. N. *et al.* (2009) 'Energy landscapes for shells assembled from pentagonal and hexagonal pyramids', *Physical Chemistry Chemical Physics*, 11(12), pp. 2098–2104.

Frank, S. *et al.* (2013) 'Bacterial microcompartments moving into a synthetic biological world', *Journal of Biotechnology*, 163(2), pp. 273–279.

Havemann, G. D. and Bobik, T. A. (2003) 'Protein content of polyhedral organelles involved in coenzyme B_{12}-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar typhimurium LT2', *Journal of Bacteriology*, 185(17), pp. 5086–5095.

Liu, Y, He, X, Lim, W, Mueller, J, Lawrie, J, Kramer, L, Guo, J & Niu, W (2018) 'Deciphering molecular details in the assembly of alpha-type carboxysome', *Scientific Reports*, 8(1), 15062.

Heinhorst, S. *et al.* (2009) 'The Pentameric Vertex Proteins Are Necessary for the Icosahedral Carboxysome Shell to Function as a CO_2 Leakage Barrier', *PLoS ONE*, 4(10), p. e7521.

Jakobson, C. M. and Tullman-Ercek, D. (2016) 'Dumpster Diving in the Gut: Bacterial Microcompartments as Part of a Host-Associated Lifestyle', *PLOS Pathogens*, 12(5), p. e1005558.

JETER, R. M. (1990) 'cobalamin-dependent 1,2-propanediol utilization by *salmonella typhimurium*', *Journal of General Microbiology*, 136, p. 887–896.

Kelley, L. A. *et al*. (2015) 'The Phyre2 web portal for protein modeling, prediction and analysis', *Nature Protocols*, 10(6), pp. 845–858.

Kerfeld, C. A. *et al.* (2005) 'Microbiology: Protein structures forming the shell of primitive bacterial organelles', *Science*, 309(5736), pp. 936–938.

Kerfeld, C. A. *et al.* (2018) 'Bacterial microcompartments', *Nature Reviews Microbiology*, 16(5), pp. 277–290.

Kinney, J. N., Axen, S. D. and Kerfeld, C. A. (2011) 'Comparative analysis of carboxysome shell proteins', in *Photosynthesis Research*, 109(1-3), pp. 21–32.

Kofoid, E. *et al.* (1999) 'The 17-gene ethanolamine (eut) operon of *Salmonella typhimurium* encodes five homologues of carboxysome shell proteins', *Journal of Bacteriology*, 181(17), pp. 5317–5329.

Laemmli, U.K. (1970) 'Cleavage of structural proteins during the assembly of the head of bacteriophage T4', *Nature*, 227, 680-5.

Laganowsky, A. *et al.* (2012) 'Atomic view of a toxic amyloid small oligomer', *Science*, 335(6073), pp. 1228–1231.

Lassila, J. K. *et al.* (2014) 'Assembly of robust bacterial microcompartment shells using building blocks from an organelle of unknown function', *Journal of Molecular Biology*, 426(11), pp. 2217–2228.

Latouf, W. G. *et al.* (2015) 'Genome and Transcriptome of *Clostridium phytofermentans*, Catalyst for the Direct Conversion of Plant Feedstocks to Fuels', *PLOS ONE*, 10(6), p. e0118285.

Liu, Y. *et al.* (2007) 'PduL is an evolutionarily distinct phosphotransacylase involved in B_{12}-dependent 1,2-propanediol degradation by *Salmonella enterica* serovar typhimurium LT2', *Journal of Bacteriology*, 189(5), pp. 1589–1596.

Ludwig, M., Sültemeyer, D. and Price, G. D. (2000) 'Isolation of CcmKLMN genes from the marine cyanobacterium, *synechococcus* sp. PCC7002 (Cyanophyceae), and evidence that CcmM is essensial for carboxysome assembly[1]', *J. Phycol*, 36, 1109–1118.

Mayer, M. J. *et al.* (2016) 'Effect of bio-engineering on size, shape, composition and rigidity of bacterial microcompartments', *Scientific Reports*, 6. p. 36899. ISSN 2045-2322.

Mcgoldrick, H.M. *et al*. (2005) 'Identification and characterization of a novel vitamin B_{12} (cobalamin) biosynthetic enzyme (CobZ) from *Rhodobacter capsulatus*, containing flavin, heme, and Fe-S cofactors*', *J Biol Chem*, 280, 1086-94.

Menon, B. B. *et al.* (2008) '*Halothiobacillus neapolitanus* carboxysomes sequester heterologous and chimeric RubisCO species', *PLoS ONE*, 3(10).

Miller, A.G., Turpin, D.H., and Canvin, D.T. (1984) 'Growth and Photosynthesis of the Cyanobacterium *Synechococcus leopoliensis* in HCO3−-Limited Chemostats', *Plant Physiology*, 75, 1064-1070.

Mori, K. *et al.* (2004) 'Identification of a reactivating factor for adenosylcobalamin-dependent ethanolamine ammonia lyase', *Journal of Bacteriology*, 186(20), pp. 6845–6854.

Obradors, N., Badia, J., Baldoma, L., and Aguilar, J (1988) 'Anaerobic Metabolism of the L-Rhamnose Fermentation Product 1,2-Propanediol in Salmonella typhimurium', *Journal of Bacteriology*, 170(5), pp. 2159-2162.

Palacios, S., Starai, V. J. and Escalante-Semerena, J. C. (2003) 'Propionyl coenzyme A is a common intermediate in the 1,2-propanediol and propionate catabolic pathways needed for expression of the prpBCDE operon during growth of *Salmonella enterica* on 1,2-propanediol', *Journal of Bacteriology*, 185(9), pp. 2802–2810.

Pang, A. et al. (2014) 'Structural insights into higher order assembly and function of the bacterial micro compartment protein PduA', J*ournal of Biological Chemistry*, 289(32), pp. 22377–22384.

Pang, A., Warren, M. J. and Pickersgill, R. W. (2011) 'Structure of PduT, a trimeric bacterial microcompartment protein with a 4Fe-4S cluster-binding site', *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, 67(2), pp. 91–96.

Parsons, J. B. et al. (2008) 'Biochemical and structural insights into bacterial organelle form and biogenesis', *Journal of Biological Chemistry*, 283(21), pp. 14366–14375.

Parsons, J. B. *et al.* (2010) 'Synthesis of Empty Bacterial Microcompartments, Directed Organelle Protein Incorporation, and Evidence of Filament-Associated Organelle Movement', *Molecular Cell*, 38(2), pp. 305–315.

Penrod, J. T. and Roth, J. R. (2006) 'Conserving a volatile metabolite: A role for carboxysome-like organelles in *Salmonella enterica*', *Journal of Bacteriology*, 188(8), pp. 2865–2874.

Petit, E. *et al.* (2013) 'Involvement of a Bacterial Microcompartment in the Metabolism of Fucose and Rhamnose by *Clostridium phytofermentans*', *PLoS ONE*, 8(1), pp. 1–12.

Price-Carter, M. *et al.* (2001) 'The alternative electron acceptor tetrathionate supports B_{12}-dependent anaerobic growth of *Salmonella enterica* serovar typhimurium on ethanolamine or 1,2-propanediol', *Journal of Bacteriology*, 183(8), pp. 2463–2475.

Price, G.D., Badger, M.R., Woodger, F.J., and Long, B.M. (2008) 'Advances in understanding the cyanobacterial CO_2-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants', *J. Exp. Bot*, 59, 1441–1461.

Rae, B. D. *et al.* (2012) 'Structural determinants of the outer shell of β-carboxysomes in *Synechococcus elongatus* PCC 7942: Roles for CcmK2, K3-K4, CcmO, and CcmL', *PLoS ONE*, 7(8), p. e43871.

Rae, B. D. *et al.* (2013) 'Cyanobacterial carboxysomes: Microcompartments that facilitate CO_2 fixation', *Journal of Molecular Microbiology and Biotechnology*, 23(4–5), pp. 300–307.

Robert Tabita, F. (1999) 'Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A different perspective', *Photosynthesis Research*, 60, pp. 1–28.

Rondon, M. R., Kazmierczak, R. and Escalante-Semerena, J. C. (1995) 'Glutathione is required for maximal transcription of the cobalamin biosynthetic and 1,2-propanediol utilization (cob/pdu) regulon and for the catabolism of ethanolamine, 1,2-propanediol, and propionate in *Salmonella typhimurium* LT2', *Journal of Bacteriology*, 177(19), pp. 5434–5439.

Roth, J. R. *et al.* (2016) 'Propanediol utilization genes (pdu) of *Salmonella typhimurium*: three genes for the propanediol dehydratase.', *Journal of Bacteriology*, 179(21), pp. 6633–6639.

Roth, J. R., Lawrence, J. G. and Bobik, T. A. (1996) 'Cobalamin (Coenzyme B_{12}): Synthesis and Biological Significance', *Annual review of biochemistry*,

50, pp. 137–181.

Sambrook, J., and Gething M.J. (1989) ‘Protein structure. Chaperones, paperones’, *Nature,* 342, 224-5.

Sambrook, J., Fitsch, E.F., and Maniatis, T. (1989) ‘Molecular cloning: a laboratory manual’, Cold Spring Harbor, N.Y., Cold Spring Harbor Press.

Sampson, E.M., and Bobik, T.A. (2008) ‘Microcompartments for B_{12}-Dependent 1,2-Propanediol Degradation Provide Protection from DNA and Cellular Damage by a Reactive Metabolic Intermediate’, *J Bacteriol* 190, 2966–2971.

Sampson, E. M. and Bobik, T. A. (2008) ‘Microcompartments for B_{12}-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate’, *Journal of Bacteriology*, 190(8), pp. 2966–2971.

Sampson, E. M., Johnson, C. L. V and Bobik, T. A. (2005) ‘Biochemical evidence that the pduS gene encodes a bifunctional cobalamin reductase’, *Microbiology*, 151(4), pp. 1169–1177.

Sargent, F. *et al.* (2013) ‘A synthetic system for expression of components of a bacterial microcompartment’, *Microbiology (United Kingdom)*, 159(PART11), pp. 2427–2436.

Sawaya, M. R. *et al.* (2015) ‘Selective molecular transport through the protein shell of a bacterial microcompartment organelle’, *Proceedings of the National Academy of Sciences*, 112(10), pp. 2990–2995.

Shisler, K. A. and Broderick, J. B. (2014) ‘Glycyl radical activating enzymes: Structure, mechanism, and substrate interactions’, Archives of Biochemistry and Biophysics, 546, pp. 64–71.

Shively, J. M. *et al.* (1973) ‘Functional organelles in prokaryotes: Polyhedral inclusions (carboxysomes) of *thiobacillus neapolitanus*’, *Science*, 182(4112), pp. 584–586.

Stojiljkovic, I., Baumler, A. J. and Heffron, F. (1995) ‘Ethanolamine utilization in *Salmonella typhimurium*: Nucleotide sequence, protein expression, and mutational analysis of the cchA cchB eutE eutJ eutG eutH gene cluster’, *Journal of Bacteriology*, 177(5), pp. 1357–1366.

Studier, F.W. *et al*. (1990) ‘Use of T7 RNA polymerase to direct expression of cloned genes’, *Methods Enzymol*, 185, p. 60-89.

Sturms, R. *et al.* (2015) 'In *Salmonella enterica*, Ethanolamine Utilization Is Repressed by 1,2-Propanediol To Prevent Detrimental Mixing of Components of Two Different Bacterial Microcompartments', *Journal of Bacteriology*, 197(14), pp. 2412–2421.

Sutter, M. *et al.* (2013) 'Two new high-resolution crystal structures of carboxysome pentamer proteins reveal high structural conservation of CcmL orthologs among distantly related cyanobacterial species', *Photosynthesis Research*, 118(1–2), pp. 9–16.

Sutter, M. *et al.* (2017) 'Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell', *Science*, 356(6344), pp. 1293–1297.

Svetlitchnyi, V. A. *et al.* (2013) 'Single-step ethanol production from lignocellulose using novel extremely thermophilic bacteria', *Biotechnology for Biofuels*, 6, 31.

Tripp, H.J. *et al.* (2010) 'Metabolic streamlining in an openocean nitrogen-fixing cyanobacterium', *Nature*, 464, 90–94.

Truan, G. *et al.* (2017) 'Spontaneous non-canonical assembly of CcmK hexameric components from β-carboxysome shells of cyanobacteria', *PLOS ONE*, 12(9), p. e0185109.

Tsoy, O., Ravcheev, D. and Mushegian, A. (2009) 'Comparative genomics of ethanolamine utilization', *Journal of Bacteriology*, 191(23), pp. 7157–7164.

Uddin, I. *et al.* (2018) 'A Generic Self-Assembly Process in Microcompartments and Synthetic Protein Nanotubes', *Small*, 14(19). p. 1704020. ISSN 1613-6810.

Wales, D. J., Clary, D. C. and Schön, J. C. (2005) 'The energy landscape as a unifying theme in molecular science', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 363(1827), pp. 357–377.

Warburg, O. and Christian, W. (1941) 'Isolation and crystallisation of enolase', *Biochem. Z*, 310, 384-421.

Warnick, T. A., Methé, B. A. and Leschine, S. B. (2002) '*Clostridium phytofermentans sp*. nov., a cellulolytic mesophile from forest soil', *International Journal of Systematic and Evolutionary Microbiology*, 52(4), pp. 1155–1160.

Yeates, T. O. *et al.* (2010) 'Structural Insight into the Mechanisms of Transport across the *Salmonella enterica* Pdu Microcompartment Shell', *Journal of Biological Chemistry*, 285(48), pp. 37838–37846.

Yeates, T. O., Crowley, C. S. and Tanaka, S. (2010) ‘Bacterial Microcompartment Organelles: Protein Shell Structure and Evolution’, *Annual Review of Biophysics*, 39(1), pp. 185–205.

Zarzycki, J., Erbilgin, O. and Kerfeld, C. A. (2015) ‘Bioinformatic Characterization of Glycyl Radical Enzyme-Associated Bacterial Microcompartments’, *Applied and Environmental Microbiology*. American Society for Microbiology, 81(24), pp. 8315–8329.